METHOD AND SOFTWARE FOR PREDICTING EMERGENCY DEPARTMENT DISPOSITION IN PEDIATRIC ASTHMA

A Thesis
Presented to
The Academic Faculty

by

Vikas R. Kumar

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Computational Science and Engineering

Georgia Institute of Technology
May 2015

METHOD AND SOFTWARE FOR PREDICTING EMERGENCY DEPARTMENT DISPOSITION IN PEDIATRIC ASTHMA

Approved by:

Professor Rahul C. Basole,
Committee Chair
School of Interactive Computing
Georgia Institute of Technology

Professor Mark Braunstein
School of Interactive Computing
Georgia Institute of Technology

Professor Nicoleta Serban
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Date Approved: 15 April 2015

To Non-Clinician MDs Everywhere

ACKNOWLEDGEMENTS

I wish to thank my faculty advisor, Dr. Rahul Basole, and my thesis committee members, Dr. Mark Braunstein and Dr. Nicoleta Serban, for their invaluable support and mentorship during this study. I also wish to thank Dr. Polo Chau. I would like to thank all collaborators on the studies I have been involved in while working for Dr. Basole, including Hyunwoo Park, Brian Minsuk Kahng, Acar Tamersoy, Mayank Gupta, Robert Pienta, and Shanshan Cao. I would like to thank the clinicians and hospital administrators who have overseen our work, including Dr. Burton Lesnick, Dr. Daniel Hirsh, Dr. Beth Schissel, James Bost, and Michael Thompson. Thank you to Matt Miller for providing us with the data and Paul Diedrich and Matt Sanders for performing file transfer requests and maintaining the remote computer. I would like to thank my professors, TAs, friends and classmates at Georgia Tech who have supported me in my classes. Last but not least, I would like to thank my mom, dad, and sister for their unwavering support as I pursue this degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

An important application of predictive data mining in clinical medicine is predicting the disposition of patients being seen in the emergency department (ED); such prediction could lead to increased efficiency of our healthcare system. A number of tools have emerged in recent years that use machine learning methods to predict whether patients will be admitted or discharged; however, such models are often limited in that they rely on specialized knowledge, are not optimal, use predictors that are unavailable early in the patient visit, and require memorization of clinical rules and scoring systems.

The goal of this study is to develop an effective and practical clinical tool for identifying asthma patients that will be admitted to the hospital. In contrast to existing tools, the model of this study relies on routine knowledge collected early during the patient visit. While most tools specific to asthma are developed using only a few hundred patients, in this study the records of 9,000+ children seen across two major metropolitan emergency departments for asthma exacerbations are used. An unprecedented amount of 70 variables is assessed for predictive strength and early availability; a novel sequence of methods including lasso regularized logistic regression and a modified "best subset" approach is then used to select the final 4-variable model. A web-application is then developed that calculates an admission probability score based on the patient parameters at the point-of-care. The methods and results of this study will be useful for those aiming to develop similar tools as well as ED providers caring for asthma patients.

CHAPTER I

INTRODUCTION

In this chapter a brief background of the current study is given, first broadly and then focusing in on the specific problem at hand. A gap is then identified which this study attempts to address in its subsequent chapters.

1.1 The U.S. Healthcare System

Currently the healthcare system of the United States is in a looming state of crisis. Annual healthcare expenditures exceed 15% of the nation's total gross domestic product (GDP); this proportion far exceeds that of other developed countries and is expected to rise to 30% by the year 2040 [20]. At the same time, the performance of our healthcare system as measured by several key indicators (e.g.,life expectancy, mortality rate) is just average compared to these countries [5]. Because the primary goal of any healthcare system is arguably to achieve the best outcome at the lowest cost, it is clear that the U.S. healthcare system needs significant overhaul and change to accomplish this more effectively.

Fortunately, steps are being taken to address the issue of inefficient U.S. healthcare. In 2009, the U.S. Congress passed the Health Information Technology for Economic and Clinical Health (HITECH) Act which calls for incentive payments for U.S. healthcare organizations that adopt electronic medical record (EMR) systems and use them to improve the quality and effectiveness of care. It is estimated that by 2019 80% of hospitals will have adopted EMR systems [25]. This will result in an unprecedented volume of clinical information being available for subsequent analysis in efforts to cut costs and improve outcomes. This leads to the critical question of how such analyses will take place.

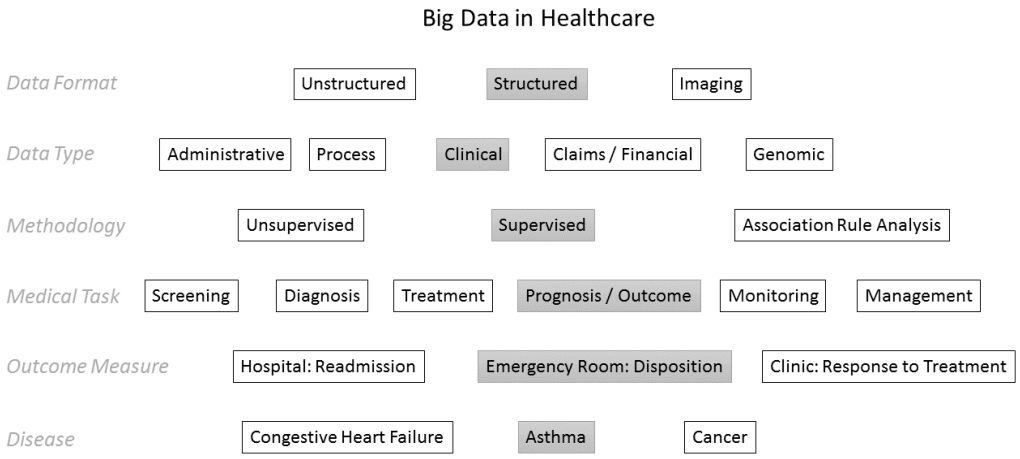

Figure 1: Schematic Overview of Healthcare Analytics Research

1.2 Big Data in Healthcare

The increased availability of electronic data in healthcare and other industries and the movement to gain knowledge from this data is often referred to as "Big Data," "Analytics," "Knowledge Discovery in Databases (KDD)," or "Data Mining" [58]. In the healthcare industry, the hope is that this data in combination with machine learning analysis techniques will help propel us into an era of improved outcomes, lower costs, and technological innovation in healthcare such as increased care automation.

Big data in healthcare is a very broad topic. To focus in more specifically, it helps to visualize various dimensions by which big data in healthcare can be classified (Figure 1). This study is concerned with only a subset of these categorizations (Figure 1, gray boxes), or specifically, outcome prediction in the emergency department for pediatric asthma, with which the remainder of this introductory chapter will be primarily concerned. An overview of healthcare analytics research that covers all of the dimensions listed in Figure 1, as well as the rationale for studying them, is given in Chapter 2.

1.3 Outcome Prediction in the Emergency Department

Now that we have skipped to a specific point of interest in healthcare big data (which will be covered in Chapter 2), next it is important to consider the general problem of predicting patient disposition in the ED, irregardless of a specific disease; such prediction could lead to improved resource planning, decreased ED waiting times, and an overall increase in the efficiency of our healthcare system. A number of models and tools have emerged in recent years that use machine learning methods to predict whether patients will be admitted or discharged; many of these studies have achieved excellent discrimination in the general ED population [21, 68].

Even the best existing tools, however, usually suffer from an important limitation: they usually require memorization of scoring systems and clinical rules for use. Easily accessible score calculators that take patient parameters as input and produce admission probability scores as output are difficult to find. However, with the increasing availability of the world-wide-web at the point-of-care through computers and mobile devices, one might expect such applications to exist already.

1.4 Asthma

When considering outcome prediction models for specific diseases, additional limitations become apparent. The focus of the current study is pediatric asthma. Acute asthma in children is an incurable disease characterized by chronic, inflammatory airway hyperresponsiveness that improves spontaneously or with pharmacologic treatment [51]. Signs and symptoms include wheezing, shortness of breath, cough, and increased respiratory rate. It is a highly prevalent disease, affecting 6.8 million children in the United States [3]. A more detailed overview of asthma is given in 2.3.1.

1.5 Outcome Prediction in the Emergency Department for Pediatric Asthma

Children experiencing asthma exacerbations often present to the emergency department (ED); in fact, in 2010 pediatric asthma accounted for 640,000 emergency room visits among children 15 years or younger [1]. One hundred sixty-nine children died in 2011 as a result of an asthma exacerbation, and the direct costs for asthma care exceed 50 billion dollars [1]. Determining whether children with asthma are healthy enough to be discharged or sick enough to warrant hospitalization therefore is a problem with financial as well as mortality-related implications.

Predictive models reported in the literature that apply specifically to pediatric asthma [55, 64, 65, 43, 42, 63, 41, 32, 31, 35, 56, 11, 59], in addition to the limitations discussed in Section 1.3, are often impractical and difficult to use. This is for a number of reasons. First, such models often rely on clinical assessment scores that require specialized consultation to obtain and are thus not routinely measured outside of clinical studies. Furthermore, these scores are often subjective and may have poor reproducibility or reliability among different providers. Additional problems associated with such clinical assessment scores are discussed in section 2.3.4. Finally, many models use predictor variables that are unavailable early on in the patients visit and thus cannot be used to predict disposition until the visit is nearly complete, at which point the disposition has already been determined by the care team.

1.6 Primary Aim of This Study

The aim of the current study is to develop an effective clinical model that can be used practically and effortlessly for identifying pediatric asthma patients that will be admitted to the hospital. Data from 9,000+ pediatric asthma patients seen at two metropolitan emergency departments is retrospectively examined across various healthcare domains including administrative, laboratory, and medication-related

domains. Over seventy predictor variables were further assessed and examined for predictive strength, objectivity and early availability; hundreds of logistic regression models were then made and tested using combinations of the best predictor variables to arrive at the final 4-variable model. A number of alternative modeling methodologies were explored. A web-application that providers can use at the point-of-care that calculates an admission probability score with minimal effort is then developed. It is hoped that the methodology and results of this study will be useful to those aiming to develop similarly practical models as well as ED providers caring for asthma patients.

The structure of this paper is as follows: Chapter 2 reviews the previous work that has been done in the area of healthcare analytics, both generally and specifically as applied to pediatric asthma. Chapter 3 discusses in detail the methods used in this study; Chapter 4 reviews the main results; Chapter 5 provides further discussion of the results and their implications and limitations; Chapter 6 describes the implementation of the web application developed as part of this study; and Chapter 7 concludes the study.

CHAPTER II

BACKGROUND

2.1 Big Data in Healthcare

As discussed in Chapter 1, data mining in healthcare is a broad research area. In this section the previously done work in various subareas of this field (see Figure 1) will be described.

2.1.1 Data Format

Healthcare data comes in many formats. The data that has been most widely used to date is structured data in the form of tables and columns. The major reason for this is convenience; the columnar organization of tabular data corresponds nicely to machine learning algorithms, which require features on which to perform clustering or make predictions. Most of the reviews of healthcare analytics research focus on this type of data [30, 15, 44].

However, unstructured or free-text data is also quite common in healthcare. For example many clinical notes written by members of the care team are written as free-text. This data is particularly important, because many clinical concepts cannot be expressed in a structured format. However, methods for mining and interpreting this data effectively for use in machine learning algorithms have yet to be fully developed. Several reviews have been written on this area of healthcare data mining [38, 62].

This study is concerned with structured data; in other words, data presented in the form of tables or in categorical formats.

2.1.2 Data Type

Structured healthcare data can be further categorized by its content, into clinical, process, financial, and genomic data. Clinical data is data related to the care of the patient, including diagnoses, medications, lab results, and physical assessments; it is the most relevant data for clinical decision-making and clinical decisions support. Because it is so broad, clinical data is often aggregated from several departments within a care organization, including the pharmacy, lab, etc.

Process data tends to be time-oriented; it typically includes the timestamps at which various administrative events, lab tests, and medication administrations occur. A separate area of research known as *process mining* is focused on the use of process data [69]; several process mining studies have been performed in healthcare [49, 50].

Financial data includes information about the charges or costs of medical activities. Such data is often provided by insurance companies and is in the form of claims data. While claims data can be useful for analyzing and comparing costs of various clinical activities, it is usually limited by the fact that usually only the dates of clinical activities are recorded, not the time of day at which they took place. It is therefore difficult to reconstruct sequences of clinical events using claims data. Additionally, results of lab or diagnostic tests may not be considered.

Genomic data includes information about the genotypes and phenotypes of patients, either obtained through genetic testing or inferred from EMR data. Because genetic tests are not routinely performed, obtaining such data is often difficult. Analysis of genomic data falls under a broad research field known as bioinformatics; a full review is beyond the scope of this study.

The current study is primarily concerned with a combination of clinical and process data.

2.1.3 Methodology

Various machine learning methodologies have been used to derive knowledge from healthcare data. There are three main categories of methods that have been used in healthcare: unsupervised learning, supervised learning, and association rule analysis.

2.1.3.1 Unsupervised Learning

The primary goal of *unsupervised learning* is simply to describe the data. It is often performed when no outcome variable (or response variable, in machine learning terminology) is used or available. Clustering is the most popular unsupervised learning method because it can be used to divide the data into different categories that may be meaningful. Specific clustering algorithms include K-means clustering and hierarchical clustering. Clustering in healthcare is often applied to the problem of patient similarity: determining which patients are most similar to each other. Several reviews have been written on the topic of patient similarity.

2.1.3.2 Supervised Learning

Supervised learning differs from unsupervised learning in that outcome variables are there to guide the learning process [34]. The model uses the outcome data for the training set to "remember" the answer when it is tested with data containing no outcome variables. Supervised learning methods are most important when it comes to prediction tasks that predict outcome based on a data sample.

Supervised learning can be divided into two flavors depending on the type of outcome variable. If the outcome variable is continuous and the goal is to predict its exact value, the learning task is called regression. If the outcome variable is categorical (e.g., telling apart apples from oranges), then the learning task is called classification.

There are several popular techniques that are frequently used to perform classification tasks in healthcare. These have been reviewed extensively elsewhere, including

generally in popular machine learning textbooks [34, 19] and clinically in review papers [15, 30]. Here we provide a brief overview of six of these techniques.

Decision Trees. A decision tree is a recursive, tree-based algorithm that classifies a data point based on splitting of its feature space. The tree is made by using principles of information theory to select the attribute and the splitting criterion that maximizes information gain at each step. A simple example of a decision tree would be to classify those at risk of heart attack by first deciding if their age was past a certain cut-off point (let's say 60 years old). If the age is less than 60, the next node in the tree checks for a systolic blood pressure of 180. A positive/negative answer to the second question results in classification of high/low risk of heart attack, respectively. On the other hand, if the age is greater than 60, the data point may then be split based on the cholesterol value, with again the exact classification dependent on the response. Decision trees are popular in medicine because they mimic the way "a doctor thinks"; however they may suffer from tree-balance problems and sensitivity to missing data.

Logistic Regression. Logistic regression is a statistical method that uses a linear equation involving predictor variables and their coefficients to calculate the response or output variable. It is a method adapted from linear regression that uses a logit transformation to restrict the output between 0 and 1. Outputs above/below a certain threshold value are classified into different categories. Coefficients are calculated using a derived equation ([34]; also see Equation 2). Logistic regression models are also popular in medicine because of the ease by which they can model binary responses (admit/discharge, live/die, etc.). Models can be adjusted when there are more than two possible response values. In more advanced models, interaction terms may be used. Picking the right combination of predictor variables to optimize model performance can be a laborious process (as the present study demonstrates).

Artificial Neural Networks. An artificial neural network uses linear combinations of predictor variables as features and then performs a nonlinear transformation on those features to perform classification. Neural networks are one of the earliest developed machine learning methods and can achieve high predictive accuracy in medicine due to their ability to model complex, nonlinear relationships. However, the increased accuracy often comes with a high computational cost, a high sensitivity to network parameters, a high potential for overfitting, and models that may be intuitively hard to interpret.

Support Vector Machines (SVMs). A support vector machine is a classifier that takes the original features of the input data and transforms them so that the classes are linearly separable; it then finds the optimal hyperplane that separates the two classes with the maximum margin. Support vector machines are a newer class of algorithms and, like artificial neural networks, are among the most accurate algorithms in machine learning [15]. The binary model can be extended for the multiclass case. However, in most cases SVMs are non-intuitive to interpret.

Naive Bayes Classifier. The naive Bayes classifier is among the simplest classifiers used; it is limited by its inability to handle nonlinear interactions [15]. The classifier assigns probabilities to class labels beginning with the prior probability which simply reflects the relative frequency ratio between the two classes; it then updates the probabilities as individual attribute values become known. Naive Bayes Classifiers may be favored in certain situations because they are robust to missing values; however, results may not be as accurate as those for SVMs, artificial neural networks and other machine learning algorithms.

K-Nearest Neighbors. The k-nearest neighbors technique, when faced with a classification task, searches the training data for the k most similar observations and classifies the new observation based on the dominant class in the k cases. It is intuitively easy to explain and interpret, as many physicians operate similarly by

basing their opinions on previous similar cases. However, it requires searching for the most similar points when performing classification, which is time consuming. It is also highly dependent on the choice of distance metric used to define closeness in the feature space.

2.1.3.3 Association Rule Analysis

Association rule analysis, or sequence mining is a method for discovering IF-THEN relationships from large amounts of sequential data [30]. Early sequence mining algorithms are defined by the Apriori algorithm which has been adapted for healthcare by substituting clinical events for purchased items and forming rules linking various medications, tests, and other clinical events to various outcomes. For example, in one study the data of more than 30,000 ICU patients was mined for associations between prolonged ICU stays and various comorbidities and medications [23]; in another study the records of over 600 patients with coronary artery disease were mined for associations between risk factors such as age, smoking and cholesterol level and th blockage of specific coronary arteries [53]; and in a third study association rule mining was combined with visual analytics to explore how diagnostic sequences are associated with the development of sepsis in lung disease patients [57].

The study is primarily concerned with techniques for predicting future events based on available data; thus, it uses supervised learning. More specifically it uses the classification method of logistic regression.

2.1.4 Medical Task

A recent review has identified six clinical tasks that can be studied using machine learning methods: screening, diagnosis, treatment, prognosis/outcome, monitoring, and management [30]. It should be noted that there is substantial overlap between these tasks.

2.1.4.1 Screening

Screening can be defined as detection of an underlying disease prior to onset of disease signs and symptoms. Studies that sttempt to use machine learning methods to screen patients have mainly focused on cancer for several reasons: 1) it is an important health problem; 2) there is a clear intervention in case of a positive screen; and 3) it has the potential to reduce costs [30, 71]. Examples of such studies include using mammography data to screen for breast cancer with a support vector machine or artificial neural network [60], and screening for prostate cancer using a support vector machine or artificial neural network [24].

2.1.4.2 Diagnosis

Medical diagnosis is defined as determining which disease or condition is causing someone's signs and symptoms. Classification methods are suited well for performing diagnosis. Usually classifier performance is compared to a "gold standard" which has near 100% sensitivity and specificity for disease detection. Again, machine learning studies for diagnosis have focused on cancer. Diagnosis of cerebrovascular and coronary artery disease are other common tasks that have been well-studied [30].

2.1.4.3 Treatment

Treatment is defined as the remediation of the health problem after disease detection. Machine learning studies for the treatment task tend to focus on predicting which patients will respond to particular treatments. Cancer has been the most studied disease in this regard [30].

2.1.4.4 Prognosis/Outcome

Prognosis is defined as predicting the patient's morbidity, mortality, outcome and chances of recovery. Outcome is defined as response to treatment. Prognosis measures are particularly common in cancer, where 1-year and 5-year survival rates are often

reported as percentages. In section 2.1.5, common measures for assessing outcome are reviewed. Because such measures can be both continuous and categorical, both linear regression and classification techniques are used within this task.

2.1.4.5 Monitoring

Monitoring encompasses observing how the patient's health state and disease changes over time with treatment. The bulk of machine learning studies for this task have studied the intensive care unit (ICU) setting. Many monitoring studies involve physiological measurements such as heart rate and electroencephalogram (EEG) recordings and employ time series analysis and classification algorithms to detect abnormalities (e.g. myocardial infarction, seizure).

2.1.4.6 Management

Finally, management in this context refers to resource allocation and scheduling. Examples of such studies include evaluation of scheduling systems, donor identification, and predicting which patients are likely to incur high operative costs.

2.1.5 Outcome Measure

In this study the aim is to predict the disposition of patients entering the ED, as opposed to screening, diagnosis, or other medical tasks. However, outcome measures vary depending on the care setting. For example, physicians in an inpatient setting may want to know which heart failure patients are likely to be readmitted in the next 30 days. Surgeons may be interested to know which of their patients have the best and worse chance for survival post-operatively. Finally, outpatient physicians may be interested to know which treatments will alleviate a patient's headaches. In the current study, patient disposition is the focus because it is an important outcome measure for the emergency department, as it has repercussions for resource allocation and staff scheduling throughout the hospital.

2.1.6 Disease

Finally, analytic studies in healthcare usually focus on a specific disease. Cancer and heart disease typically predominate. In the current study pediatric asthma is the focus; for an overview of asthma and its health effects on children nationwide, see Section 1.4.

Now that a brief review of data mining studies in healthcare has been completed, studies specific to the problem at hand (outcome prediction in the ED for both the general population and specifically for pediatric asthma patients) will next be reviewed.

2.2 Outcome Prediction in the Emergency Department

Many studies over the past twenty years have attempted to predict hospital admission in the emergency department. Some of these studies have used more subjective measures as predictors, such as the intuition of paramedics or triage nurses. Some have focused on specific diseases such as chronic obstructive pulmonary disorder (COPD) and heart attack (myocardial infarction, MI). Lastly, many studies looked at data obtained during the entire duration of the ED visit and so may not apply to the early identification of admitted patients. Models designed specifically for pediatric asthma will be discussed in Section 2.4.

Logistic regression was the most common classification used in these studies; however there was a 2006 study that achieved high accuracy using an artificial neural network [47]. That study used data from 43,077 patients; the final model included nine variables related to age, gender, ICD-9 diagnosis, and arrival method. The model also included information about the presence of certain lab tests, x-rays, and electrocardiogram (EKG) tests that occur late in the ED visit. The model obtained an AUC of 0.897. A 2007 study [28] examined 47,933 ED visits and used a Bayesian network to make predictions; however, their AUC was only 0.833.

More recently, studies have emerged that attempt to apply prediction algorithms to the general ED patient population early in the visit using logistic regression [52, 68, 21]. A 2008 study from the U.S. [52] used data from 401 patients obtained from two EDs to make a logistic regression model that included the presence of certain symptoms and comorbidities as well as age. Their model obtained an AUC of 0.80, and specificity at the selected cutoff was 69%. A 2011 study from Singapore [68] retrospectively examined 317,581 cases for variables including demographics, acuity category, and active diagnoses. They found that admission rates depended on age, ethnic group, arrival mode, acuity status, past ED visits/admission, and the presence of several chronic conditions such as hypertension, diabetes, and dyslipidemia. They used stepwise logistic regression to make a predictive model that included all of the significant predictors, and they obtained an AUC of 0.849. A 2015 study from Australia [21] retrospectively examined over 320,000 cases across different centers for very similar variables (age, triage acuity, physiological early warning score, arrival method, referral source, and history of previous admission). A logistic regression model using those six variables achieved an AUC of 0.8774. In addition to the model, they constructed an admission prediction score that can be used at the bedside.

2.3 Asthma

In this section an overview of asthma is given and current methods for measuring asthma severity are reviewed. The historical need for a simple type of "scoring system" for ED use becomes apparent, and the rise of those systems is reviewed. Finally, important limitations of these systems that hinder their functionality in the ED and serve as a partial motivation for the current study are described. For a brief definition of asthma and its morbidity and mortality, see Section 1.4.

2.3.1 Asthma: Overview

What follows in this section is a brief overview of asthma.

2.3.1.1 Epidemiology and Pathogenesis

While asthma can affect individuals at all ages, about one-half of cases develop before age 10 [51]. In childhood it affects males more than females; there is a 2:1 ratio.

Asthma attacks occur when certain stimuli reach the cells on respiratory airway surfaces of hypersensitive individuals [51]. There is a high concentration of immunologic cells, which have the function of protecting the airways from potentially harmful stimuli, in these areas. In response to certain stimuli (discussed in the next section), these cells release "inflammatory mediator" substances which have particular effects on the respiratory airways. The effects include constriction of the airways and hypersecretion of mucus, which in turn impede air flow and exchange. As a result, patients feel short of breath, cough frequently, and usually experience wheezing. Visible activation of the accessory muscles (chest and shoulder muscles) may also be present.

2.3.1.2 Inciting Stimuli

Stimuli that can induce asthmatic episodes include pharmacologic stimuli such as aspirin and coloring agents, environmental and air pollutants, occupational factors such as dusts and enzymes, infections, exercise, and emotional stress. Of these, respiratory infections are the most common culprit of asthma exacerbations [51].

2.3.1.3 Diagnosis

See Section 2.3.2 for an explanation of how spirometry is used to demonstrate reversible airway obstruction.

2.3.1.4 Treatment

Many drugs are designated for treating asthma attacks. Generally, they can be divided into two categories: 1) quick relief medications (including albuterol and ipratropium) and 2) long-term control medications (corticosteroids and long acting beta$_2$

agonists) [51].

2.3.2 Asthma: Methods for Assessing Severity

The textbook, gold-standard class of procedures for assessing the severity in asthma is the pulmonary function test (PFT). [54]. We will review two methods of this class here: spirometry and peak expiratory flow rates (PEFR).

Spirometry refers to the measurement of lung volume and breathing force using a spirometer. A spirometer is a large machine in which children blow and that then takes various measures about the rate and quantity of expired air. The exact measure that is the gold standard for asthma severity is the amount of air forcibly expired in one second (FEV_1). FEV_1 is often expressed as a fraction over the forced vital capacity (FVC), which is proportional to the lung capacity. In children, An FEV_1/FVC ratio that is below 85% of the predicted ratio (calculated by age and weight) and that improves by at least 12% with administration of a medication called a bronchodilator is considered diagnostic of asthma [39]. The ratio of the FEV_1 over the FEV_1 can also be used directly to monitor asthma exacerbation severity, although there are no standard normal and abnormal measurements for different age groups [18].

While spirometry certainly objective, it is rarely performed in the ED for pediatric asthma. This is due to two main reasons. First of all, the urgency and necessity of medical treatment may not allow time for children to perform spirometry, since it takes up to a half hour to perform in children [39]. Second, FEV_1 readings are dependent on proper technique which is not consistently achieved by children of certain ages (e.g., pre-schoolers) [54].

The second method, peak expiratory flow rate (PEFR) measurement, is a portable alternative to spirometry since it uses a small plastic device instead of a spirometer to gauge the amount of air expired forcibly. However, PEFR measurements are also seldom taken in emergency departments, due mainly to their high inaccuracy in

children; a recent review found incorrect use of PEFR meters by children at almost all steps of the procedure [18].

Therefore, because objective air flow measurements are not suited for the ED in children, physicians have filled this diagnostic gap by devising clinical asthma scores (used interchangeably with clinical asthma scoring systems and clinical asthma assessment scores in this study), which are calculated by estimating the severities of a few asthma signs and symptoms (wheezing, cough, respiratory rate) on 2- to 3-point scales and adding these subscores to obtain an overall score.

2.3.3 Clinical Asthma Scores: Overview

Before delving into clinical scoring systems specific to asthma, it is important to discuss clinical scoring systems in general. Previous studies have identified three major criteria that must be met before such rules can be prospectively applied in medicine: 1) the demonstration of the need for such a rule; 2) the derivation of a rule using methodologic standards; and 3) the prospective validation of a rule [13]. The first criterion requires no explanation. Methods often used to fulfill the second criterion include a clear definition of outcome and candidate predictor variables and selection of particular variables based on statistical techniques such as regression analysis.

For the third criterion, studies are often performed subsequent to score development that subject the score to a variety of validation measures [18]. These include: a) item generation, or the basis of how score components are generated; b) reliability, or the degree to which scores can be replicated by different providers; c) validity, or the degree to which scores measure what they purport to measure; d) responsiveness, or the ability of a score to detect clinically meaningful change; and e) usability, or the ease of use in practice. As predictive analytics in healthcare continues to grow in importance (see Section 1.2), perhaps a sixth feature should be added: prediction, or

the ability to which scores can foretell patient outcomes.

Clearly based on Section 2.3.2 the first criterion is fulfilled for asthma. However, as the following review will show, compliance with the second criterion is rarely achieved, and application of the third criterion often reveals deficiencies of these scoring systems.

The earliest reported asthma scoring system was introduced in 1966 and was called the Bronchiolitis Score (BS) [26]. It is a 27-point score, with $0-3$ points given to each of nine criteria based on severity: cyanosis, activity, cough, respiratory rate, retraction score, resonance, wheezing, expiration/inspiration, and liver/spleen. It was originally developed not for routine clinical use, but to test the effects of corticosteroids on the treatment of bronchiolitis in a research study. As the name suggests, it was not developed for asthma, but instead for bronchiolitis, a related but separate disease. Finally, there was no methodological standard used to select these criteria, nor was there any attempt to validate the score.

Despite this, other scores soon derived from the BS and were geared towards clinical use. In 1984 the Pulmonary Index (PI) was developed [14], which was a refinement of the BS [13]. The PI used four criteria graded on a $0-3$ point scale: respiratory rate, wheezing, expiration/inspiration, and wheezing. In this study the PI was compared with spirometry measures, and there was a significant correlation between the two. In 2002 the pulmonary score (PS) was developed [66], which was derived from and very similar to the PI, using the same scale as and three of the four criteria from the PI.

These scores serve as just one example of how clinical scores evolved from each other and had little methodological basis. By the 1990s at least one dozen clinical asthma scores had been reported in the literature, with possibly many more being used internally by different healthcare organizations. A complete review of all of these scoring systems is beyond the scope of this work; interested readers are referred to previous reviews that have been done on this topic [70, 18]. However, three asthma

scores that have been developed more recently and as a result have undergone more testing and scrutiny during the validation process will be reviewed here: the Preschool Respiratory Assessment Measure (PRAM) [22], the Pediatric Asthma Severity Score (PASS) [33], and the RAD score [12] (for a summary, see Table 1).

The 12-point PRAM was first reported in 2000 and was developed based on a stepwise linear regression of 18 clinical signs with respiratory resistance as the response variable; the five clinical signs that showed the best discriminatory power and responsiveness to treatment were included in the final score. The PRAM scoring criteria are shown in Table 1. A 2008 study [29] by the same authors subjected the PRAM score to further validation testing using measures describe earlier in this section. Their findings are shown in Table 1.

The 7-point PASS was developed and validated in 2004 and included three to five criteria selected by physicians based on their face validity. The PASS scoring criteria and validation results are shown in Table 1.

Finally, the RAD score was developed in 2011 and includes three criteria – respiratory rate, accessory muscle use, and decreased breath sounds – spread over a three-point scale. The authors framed it as a more usable alternative to the PRAM and PASS scores. In their validation they compared the RAD with the PASS and PRAM and found that the RAD performs more favorably than the other scores. Their findings have yet to be confirmed by a second party.

Table 1: Summary of Recently Developed Clinical Asthma Scoring Systems

Score	**Scale**	**Signs and Symptoms Included**	**Item Selection**	**Reliability**	**Validity**	**Responsiveness**	**Prediction**
PRAM [22]	0 – 12	1. Suprasternal Retractions; 2. Scalene Muscle Contraction; 3. Air Entry; 4. Wheezing; 5. SaO2	Stepwise Linear Regression of 18 Clinical Signs	Internal Consistency: Cronbach's $\alpha = 0.71$; Inter-rater Agreement: K = 0.78		Guyatt coefficient: 0.7; Effect size: 1.1	AUC: 0.78 – 0.86 [11]
PASS [33]	0 – 7	1. Wheezing; 2. Work of Breathing; 3. Prolongation of Expiration; 4. Abnormal Respiratory Rate	Chosen based on previous scoring systems, usability, and face validity	Inter-rater Agreement: K = 0.79 – 0.83	Correlation with SaO2: Pearson's = -0.44 – -0.31; Correlation with PEFR: Pearson's = -0.34 – -0.25	Effect size: 0.62; Change in Score from Start to End of Treatment: 51 – 79 % (Discharged); 25 – 32% (Admitted)	AUC: 0.84 – 0.88 [33, 32]
RAD [12]	0 – 3	1. Respiratory Rate; 2. Accessory Muscle Use; 3. Decreased Breath Sounds	Chosen based on construct validity and usability		Correlation with %FEV1: $R^2 = 0.426$	Correlation with %Δ FEV1: $R^2 = 0.139$	

2.3.4 Clinical Asthma Scores: Limitations

Currently, clinical asthma scores are the cornerstone for asthma severity assessment and monitoring in the ED. They are more practical to perform than more objective measures such as spirometry and PEFR testing. So why is this problematic? Returning to the three criterion discussed in Section 2.3.3 for successful development of clinical scoring systems, it is clear that these scores fail to meet the second criterion of having a methodological basis, and their validation (the third criterion), when performed, reveal the deficiencies of these scores, as will be outlined here. For an overview of how clinical scoring systems are commonly evaluated, see Section 2.3.3.

2.3.4.1 Nonsystematic Item Selection

For scores to be optimally valid, there should be a sound basis for item selection. Most clinical asthma scores, however, seem to be based simply on the opinions of a few physicians. As reviewed in Section 2.3.3, the Bronchiolitis Score, which is the precursor of many modern scoring systems including the Pulmonary Score and the Pulmonary Index, was developed based on the "intuition" of three physicians. Furthermore, of the three recently developed scoring systems (the PRAM, PASS and RAD), only one of them (the PRAM) was developed using a statistical technique (stepwise linear regression); the other two were again developed based on the opinions of a few physicians.

2.3.4.2 Suboptimal Reliability

The most valid scoring systems have a high degree of reliability, also known as reproducibility between different providers. However, given the subjective way in which many of the score items are assessed, it is no surprise that scores taken by different care providers are often different. For example, one of the criterion for the PRAM is "air entry." The patient receives zero points if air entry is "normal," one point if "decreased at bases," two points if there is a "widespread decrease", and four points

if air entry is "absent or minimal" [22]. Assessing the air entry involves use of a stethoscope and listening to lung sounds. Different stethoscopes may have different auditory properties, and the users may also perceive lung sounds differently. The rating of a medical student or resident, who are relatively inexperienced, may differ from that of a more experienced physician. Other items used in the PRAM, including wheezing and the degree of muscle contraction, are similarly ambiguous. Therefore it is no surprise to see that the weighted K statistic, used to assess inter-rater agreement, is 0.78. The authors say that a weighted K statistic above 0.7 is "good." Can we do better? The authors also call for an independent confirmation of their findings, which has yet to be performed.

2.3.4.3 Suboptimal Validity

Validity, or the degree to which a scoring system measures what it aims to measure, is separate from validation; while validation refers to the entire scoring system certification process, ensuring validity is just one aspect of that. The purpose of asthma scoring systems is to quantify asthma severity; a very severe asthma attack should therefore be correlated with a very low clinical score, if the scoring system is valid.

There are three different types of validity [18]. Face validity is the qualitative degree to which a score measures what it's supposed to measure. Construct validity is the degree to which the measure corresponds to theoretical constructs. Finally, criterion validity is the degree to which the measurement corresponds to a gold standard. As discussed in Section 2.3.2, pulmonary function tests (PFTs) are the gold standard for assessing asthma severity.

A recent review of ten clinical asthma scoring systems found that most scores fail to even assess their own validity [18]. The PRAM, PASS, and RAD do report validity; however, the methods by which they do so and the corresponding results are problematic. The PRAM tests validity by comparing internal consistency between

the PRAM score and each of its five components [29]. Of course the PRAM will be consistent with its own subparts; a true test of construct validity would involve correlating the PRAM with an independent measure not in the PRAM such as hospital length of stay or respiratory rate (since it already includes SaO2). The PASS did correlate its score with independent measures; however, its results were subpar, finding Pearson's correlation coefficients of -0.44 to -0.31 with SaO2 and -0.34 to -0.25 with PEFR measurements [33]. The RAD tests for criterion validity by directly correlating the RAD score with FEV_1; the r^2 is only 0.426 [12].

2.3.4.4 Suboptimal Responsiveness

Responsiveness is the degree to which a score change reflects clinically important events. The PRAM score reports an effect size [72], calculated using the change in PRAM between triage and disposition, of 1.1 [29]. The PASS score reports an effect size of 0.62 and finds that the score change is 51 – 79% from the start to end of treatment [33]. The RAD correlates the score with the $\%\triangle FEV_1$ and obtains a r^2 of 0.139 [12]. The questions remains as to whether responsiveness can be further optimized by choosing different predictors.

2.3.4.5 Poor Usability

The ease of use of clinical scoring systems is a particularly important factor; if a score is time-consuming and subjective to perform it is not only inefficient but also dangerous. Many of the scores may be used for all patients while performing a research study, but if a score has poor usability it will translate to infrequent use in actual, non-research clinical care. The original Bronchiolitis Score (reviewed in Section 2.3.3) is obviously unusable in a clinical setting; rating nine items on scales of 0 – 3 may well take 10 – 15 minutes.

The PRAM, PASS and RAD don't perform any usability testing; however assessment of 4 – 5 criteria does take significant time. The RAD may be more usable than

the PRAM and PASS since it only uses three criteria, each graded on a one-point scale. However these scores in their current form also require adding subscores. If there was a way in which such scores could be calculated automatically with the use of computers, that would be ideal. No clinical asthma scoring system has reported attempting this.

2.3.4.6 Suboptimal Prediction

Scores that can predict eventual patient outcomes could potentially be useful. Predictive abilities of the PRAM and PASS are summarized in Table 1. The PASS has a particularly strong average AUC, almost 0.9 [32]. However, the PASS prediction study relies on an additional criterion: the number of albuterol treatments received in the ED. This predictor is not available early in the ED visit and may compromise effective prediction in the clinical setting. The PRAM has a "good" AUC of 0.78 – 0.86 [11].

2.4 Outcome Prediction in the Emergency Department for Pediatric Asthma

Refer to Table 2 for a review of studies that attempt to predict ED outcome in pediatric asthma.

Table 2: Summary of Outcome Prediction Studies for Pediatric Asthma

Study	**N**	**Variables Included**	**Methodology**	**Outcome Measure**	**Model Performance**
[55]	200 (100 train, 100 test)	Initial peak flow; history of treatment in 24 hrs; age at asthma onset; # prev hosp for asthma	Discriminant analysis	Disposition	Sn: 33% (train); Sn: 0% (test)
[64]	133	FVC; FEV1; cyanosis; accessory muscle use; insp breath sounds; exp wheezing; cerebral function	ANOVA; Newman-Keuls multiple comparison procedure	Clinic disposition	Sn: 71% Sp: 93%
[65]	156	Insp breath sounds; wheezing; RR	Logistic regression	Disposition	Sn: 95%; Sp: 66%
[43]	200	HR; RR; dyspnea; wheezing; accessory muscle use; pulsus paradoxus	Decision tree	Disposition	Sn: 73%; Sp: 95%
[42]	71	HR; RR; dyspnea; wheezing; accessory muscle use; pulsus paradoxus	Correlation	FEV1; SaO2	r^2: 0.52 (FEV1); r^2: 0.49 (SaO2)
[63]	120	Accessory muscle use; dyspnea; RR; SaO2; FEV1	Logistic regression	Disposition	AUC: 0.57 (baseline); AUC: 0.83 (post-treatment)
[41]	720	Altered consciousness; exhaustion; speech; pulsus paradoxus; HR; cyanosis; wheezing; peak expiratory flow; FEV1; SaO2	Descriptive analysis	Disposition	Descriptive; 1hr assessment more accurate than triage assessment

Table 2 – *Continued from previous page*

Study	**N**	**Variables Included**	**Methodology**	**Outcome Measure**	**Model Performance**
[32]	1221 (852 train, 369 test)	PASS score (wheezing, work of breathing, prolonged expiration, RR); Num albuterol treatments in ED	Logistic regression	Disposition	train: AUC 0.89, Sn 82%, Sp 84%; test: AUC 0.92, Sn 90%, Sp 80%
[31]	362 (239 train, 123 test)	Wheezing; retractions; HR; RR; Previous visit to chest clinic	M4 Decision tree; Logistic regression	Asthma exacerbation severity (Mild, Moderate/Severe)	M4: AUC 0.83, Sn 84%, Sp 71%; LR: AUC 0.74, Sn 69%, Sp 68%
[35]	2.4 million	Tachycardia; systolic hypotension; diastolic hypotension; tachypnea; abnormal temp; hypoxemia; sex; arrival by EMS; triage index; Prev visit 72 hrs; Prev discharge 1wk; insurance status	Age-specific forward-selection logistic regression	Odds of admission	OR varied greatly by age for most parameters; diastolic hypotension and tachycardia universally predictive
[56]	1641	Asthma; presence of bronchopneumonia; presence of diarrhea/dehydration/UTI	Multiple logistic regression	Odds of admission	Asthma + BP: OR = 3; Asthma + others: OR = 4.7
[11]	297	PRAM score (on admission, 2hr, 3hr, 4hr)	Logistic regression	Disposition	AUC: 0.76 (triage), 0.85 (2hr) 0.85 (3hr), 0.83 (4hr)
[59]	539	SaO2	Multivariate forward stepwise logistic regression	Disease severity (mild, severe)	AUC: 0.75

2.5 Summary

Having reviewed the background and current progress in areas related to predicting patient disposition in pediatric asthma, four points have become apparent:

1. There is a need for an effective yet practical asthma outcome prediction model. The performance of existing asthma prediction models, as measured by the area under their receiver operating characteristic (ROC) curves (AUC), could be further improved. The model that has achieved the highest AUC [32] uses information throughout the patient visit and is therefore limited for providing early predictions. Furthermore, the developed model needs to be translated into a clinically useful one by implementing either a scoring system or web application.

2. Predicting asthma disposition with objective, readily available data holds promise over existing methods. Current models for predicting asthma disposition often rely on clinical assessment scores which are subjective, difficult to perform and unreliable in many cases. Using objective, readily available information has the potential to make these models more robust and widespread.

3. Existing predictive modeling and model selection techniques may be limited or outdated. Current logistic regression studies in healthcare often choose variables using a subjective "purposeful selection" method. As data grows in volume and availability the number of predictors could also greatly increase; for example the current study had access to over 70 predictor variables. With such high numbers of variables, more systematic methods for variable selection are needed, perhaps implemented in a multi-stage, sequential fashion.

4. As EMR implementation rises, the possibility of predicting patient disposition in real-time should be explored. Doing so has implications for further automation and technological advancement of the U.S. healthcare system.

This multi-faceted study attempts to address all four of these observations:

1. Development of practical software for effectively predicting patient

disposition in asthma. The software is available at the point-of-care for both computers and mobile devices. It effectively predicts outcome using just four variables - a manageable number of predictors.

2. Demonstration that effective prediction can be achieved using objective, readily available data. The hypothesis that objective, readily available data can predict disposition on the same level as subjective data is tested and proven. All models in this study use predictors that are routinely obtainable. Furthermore, a specific model is developed using "Purposeful Selection" that uses objective predictors to achieve strong predictive results.

3. Development of new methodology based on lasso regularization and a modified "best subset" approach for logistic regression model selection. In Chapters 3 and 4 a new sequence of methods for variable selection is applied.

4. Development of a dynamic model that updates predictions and improves in real-time. The ability to improve model performance using objective, readily-available data is demonstrated.

CHAPTER III

METHODS

3.1 Data

Data was requested for all emergency department visits that occurred between January 1st, 2013 and January 31st, 2014 that had a primary diagnosis of asthma (ICD-9 code 493.xx). Egleston and Scottish Rite hospitals (both of the Children's Healthcare of Atlanta) were the two hospitals from which data was collected. The data was extracted from Children Healthcare of Atlanta's data warehouse and was received in a relational database format containing 24 tables in the form of .csv files (roughly 12 tables for each hospital). Table 3.1 summarizes the received files and their contents. Data was stripped of all protected health information (PHI); all visit and patient IDs were anonymized. Visits for which there was no emergency department exit timestamps were excluded from the analysis. This project was reviewed and approved by the Georgia Institute of Technology Institutional Review Board.

3.2 Outcome

The target outcome (or response variable) chosen for model development was the patient disposition at the end of the emergency department visit. It was a binary outcome based on the "ED Dispo" attribute in the Visits.csv data file. Patients who had a value of "Discharge" or "Discharge after orders" were assigned to the "discharged" outcome (value of zero), while patients who had a value of "Admit" (to regular ward) or "Admit to ICU" were assigned to the "admitted" outcome (value of one). Patients were excluded if they had "ED Dispo" values that differed from these four values (e.g., "Admit to Operating Room" or "Transferred to Another Facility").

Table 3: Data Files and Associated Contents

Filename (.csv)	**File Contents**
AsthmaSteroidED	Steroid administration events and their corresponding visit IDs, timestamps, administrating providers, drug names, dosages, and formulations
AsthmaSteroidED2	Steroid administration events and their corresponding visit IDs, timestamps, administrating providers, drug names, dosages, and formulations
Charges	Itemized charge information for each visit, including service, lab test, and medication charges
CRSScores	Clinical Respiratory Score (CRS) assessment events, including visit IDs, timestamps, and scores
Diagnosis	ICD-9 diagnostic codes associated with each visit
LabResult	Lab events and their corresponding visit IDs, timestamps, test names, ordering providers, and test results
Medication	Medication administration events and their corresponding visit IDs, timestamps, administrating providers, drug names, dosages, and formulations
Visit	Visit IDs and corresponding demographic information, provider information, administrative timestamps, and pre-arrival/triage data
VisitDepartments	Hospital departments occupied for each visit ID
VisitProcedures	Procedures underwent for each visit ID
VisitProviders	Providers associated with each visit ID
Vitals	Vital Sign assessment events (respiratory rate, pulse, blood pressure, temperature, and SaO2) and their corresponding timestamps and values

An alternative possible outcome considered but not chosen was the change in CRS score from the initial score timestamp to the final score timestamp ($\triangle$CRS). The main reason $\triangle$CRS was not chosen was that it required two CRS scores to be measured, while less than half of ED patients had even one score measured. Disposition, on the other hand, was recorded for every patient.

Finally, another possible outcome would have been to have a multiclass (k >2) outcome variable with three possible values: Discharge, Admission to Ward, and Admission to ICU. This option was not chosen because it would have greatly limited the possible machine learning algorithms that could be used, since most algorithms are designed for binary outcomes (see Section 2.1.3).

3.3 Data Pre-Processing

Data from the two Visit.csv files (one for each hospital) was imported into a MATLAB 2013b environment (Mathworks: Natick, MA) as a cell array object in string format, using the textscan() function. The cell array was of dimension $n \times p$, where n is the number of patients (observations) in the study and p is the number of features (attributes). Initially each Visit.csv table included approximately 145 features.

Six features from each of the two Vitals.csv files were appended to the initial cell arrays. For each vital sign the initial value was chosen in order for our model to rely on the earliest data possible.

For some of the supplementary models containing "late" predictors, information regarding the number of labs drawn and medications given by certain time points were used as predictor variables. In those cases we first filtered out lab draws and medication administrations that occurred after the patient exited the ED. Then, the relevant information was appended to the cell array.

Variables were converted to double format where appropriate, and the cell array data structure was converted to the dataset data structure to be compatible with

MATLABs new algorithms for performing univariate and multivariate logistic regression.

3.4 Candidate Variables

Of the original 150 attributes, 70 candidate variables were then selected for further analysis. The 70 variables were chosen based on clinical and intuitive judgement as to whether they could possibly be correlated with patient disposition. Examples of excluded attributes included some features, such as timestamps, that structurally could not be represented appropriately as predictor variables. Visit and patient ID numbers, as they are random, were also excluded. Features that overlapped with ones already chosen for analysis were excluded. Finally, it should be noted that age was excluded because on preliminary qualitative examination it did not appear to be a significant predictor of admission rate (Figure ?). The 70 variables included in the univariate analysis are shown in Table 4.

Table 4: Predictor Variables Included in Univariate Analysis

Demographic Variables	
Sex	Race
Language	Interpreter Needed
Financial Variables	
Primary Payor	Financial Class
Financial Class Group	
Pre-Arrival/Triage Variables	
IBEX Admit Code	ED Admission <72hrs
ED Admission >72hrs	Asthma Admission <72hrs
Asthma Admission >72hrs	IBEX Acuity
ED Visit Shift	Day of Week
IBEX Arrival Method	Arrival by EMS
Triage Acuity	
ED Assessment - Vital Signs	
Systolic Blood Pressure	Diastolic Blood Pressure
Temperature	Respiratory Rate
Heart Rate	SaO2

Table 4 – *Continued from previous page*

ED Assessment - Laboratory	
Rapid Strep Test	Basic Metabolic Panel
Comprehensive Metabolic Panel	Complete Blood Count (w/ Diff.)
Blood Culture	C-Reactive Protein
Any Labs	Lab Count
ED Assessment - Radiology	
CT: Head	CT: Abdomen
CT: All	X-ray: KUB
X-ray: Chest	X-ray: All
Radiology Count	
ED Treatment	
IV Antibiotics	IV Fluids
IV Zofran	Medication Count
Process Variables	
ED Length of Stay (hrs)	Arrival to Triage Start (mins)
Arrival to Triage End (mins)	Arrival to Room (mins)
Arrival to 1st Provider (mins)	Arrival to First Attending (mins)
Arrival to Registration (mins)	Arrival to Disposition (mins)
Arrival to ED Exit (mins)	Triage Start to Triage End (mins)
Triage End to Vital Sign Restart (mins)	Triage End to Init Assessment (mins)
Triage End to Room (mins)	Triage End to First Attending (mins)
Triage End to 1st Provider (mins)	Room to Nurse (mins)
Room to 1st Attending (mins)	Room to 1st Provider (mins)
Room to 1st Resident (mins)	Room to Disposition (mins)
1st Attending to Disposition (mins)	1st Provider to Disposition (mins)
1st Provider to Exit (mins)	Disposition to Exit (mins)

3.5 Logistic Regression: Overview

3.5.1 Linear Regression

Logistic regression is an extension of linear regression that allows for the modeling of categorical response variables. Suppose there is an input vector X of predictor variables, and one is interested in predicting the value of a continuous response variable Y. The basic linear regression model is

$$f(X) = \beta_0 + \beta_1 X \tag{1}$$

where the β's are vectors of coefficients and $f(X)$ is the prediction [34]. In regression a commonly used method to measure the error is to take the residual sum of squares of the difference between each value y and corresponding predicted value $f(x)$. This error will depend on the chosen β:

$$RSS(\beta) = \sum_{i-1}^{n} (y_i - f(x_i))^2. \tag{2}$$

Using matrix math we can find the value of β that minimizes the residual sum of squares [34]:

$$\hat{\beta} = (X^T X)^{-1} X^T \mathrm{y}. \tag{3}$$

3.5.2 Logistic Regression

While linear regression can predict a continuously-valued response, it is not optimized to model the probability that a categorically-valued response will occur (e.g.,the predictions of linear regression do not take values between 0 and 1). A transformation can be used to convert the continuous response to a value between 0 and 1, which results in the basic logistic regression model [4]:

$$\ln\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 X \tag{4}$$

where the response μ takes on a value between 0 and 1. The coefficient vectors β can be found using the same least squares approach of Equation 3.

For each predictor variable the odds ratio OR (defined as the increase in odds of observing a positive response with a one-point increase in the predictor variable) can be calculated using the following equation [67]:

$$OR = e^{\beta}. \tag{5}$$

Odds-ratio values above 0 indicate a positive correlation between the variable and the response; while values below 0 indicate the opposite.

3.5.3 Model Selection Methods

Often model performance is dependent on optimizing the predictor variables used in the model. Using too many predictors can often increase the model variance, while using too few predictors can increase the model bias.

3.5.3.1 Best-Subset Selection

Best-subset selection finds for each $k \in \{0, 1, 2, ..., p\}$ the predictor subset of size k that gives the smallest error [34].

3.5.3.2 Shrinkage Methods: The Lasso

Shrinkage methods shrink the regression coefficients by restricting the size of the coefficients and penalizing those having larger values [34]. β is selected using the equation

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}. \quad (6)$$

Note that Equation 6 is similar to the optimization problem presented in Equation 2 where an additional term is added to penalize the sum of the absolute values of the coefficients.

3.5.3.3 Univariate Logistic Regression

Univariate logistic regression can be performed on each of the predictors to find which are most strongly correlated to the outcome and therefore most suitable for inclusion into the model. The equation used to predict the response is identical to Equation 4, where β_0 and β_1 each have a length of 1.

3.5.3.4 Other Selection Methods

Other methods for subset selection include stepwise selection methods (such as forward- and backward-stepwise) in which the model is optimized by sequentially adding or removing the predictor that results in the most improved fit. These are greedy algorithms that may save on computation time but are not guaranteed to provide optimal (or even identical) solutions across different runs [34].

Ridge regression is an alternative shrinkage method in which the L_1 lasso penalty is replaced by the L_2 ridge penalty [34].

Finally, in applied logistic regression predictors may often be chosen subjectively based on knowledge of the subject at hand. In medicine such models have been termed "purposeful selection" models [36].

3.6 Univariate Logistic Regression

In order to identify variables which were significantly related to the outcome, univariate regression was performed by calling the fitglm() function, using each column number as a predictor variable in turn with the response variable set to the patient disposition. The distribution of the function was set to binomial. By default, the fitglm() function for the binomial distribution fits a generalized linear model according to Equation 4. For each coefficient, the 95% confidence intervals were calculated using the coefCI() function. Odds ratios (and 95% confidence intervals) were calculated using Equation 5. If the confidence interval for a particular variable did not include 1, that variable was deemed significant. P-values were also used to assess significance.

3.7 Intermediate Models Containing "Late" Predictors

To support the goals developed in Section 2.5, additional models were made that used "late" variables in addition to "early" predictor variables. Although these three models cannot be used early during patient visits, they may still serve some useful

purpose as will be explained.

3.7.1 "Purposefully Selected" Model

The variables included in the first multivariable model were chosen based on the univariate results. A number of factors were considered when choosing variables: 1) variables having significant p-values were favored over non-significant variables; 2) variables having very high or very low odds ratios (OR) were favored over variables having an OR near one; 3) earlier data was used over later data whenever possible; 4) variety across different data domains (e.g. financial, clinical, demographic, etc.) was encouraged since they would be less likely correlated; 5) variables having a larger scale were used over those with a smaller scale; and 6) variables having more of an equal split were favored over variables that were lopsidedly positive or negative.

Table 5: Predictor Variables Included in Supplementary Models

"Purposefully Selected" Model	
Financial Class Group	IBEX Admit Code
ED Admission <72hrs	IBEX Acuity
ED Visit Shift	Arrival by EMS
SaO2	Rapid Strep Test
Lab Test Count	Radiology Test Count
IV Fluid Order Count	Medication Count
Arrival to Triage Start (mins)	Triage End to Room (mins)
"Lasso Regularized" Model	
Financial Class	Financial Class Group
IBEX Admit Code	Systolic Blood Pressure
Diastolic Blood Pressure	Rapid Strep Test
Basic Metabolic Panel	Lab Test Count
Head CT	IV Fluid Order Count
Triage End to Initial Assessment (mins)	Disposition to Exit (mins)
"Dynamically Updated" Model	
Financial Class	Triage Acuity
Admission Source	Arrival by EMS
ED Visit Shift	ED Admission <72hrs
Asthma Admission <72hrs	Day of Week
Lab Test Count (dynamic)	Medication Count (dynamic)

Table 5 – *Continued from previous page*

Arrival to Room (mins)	Arrival to Triage Start (mins)

3.7.2 "Lasso Regularized" Model

A lasso regularized model was constructed using procedures similar to those outlined in Section 3.8.2, except that all of the predictors were included in the initial pool of variables. Variables that were selected by this lasso regularization procedure are shown in Table 5.

3.7.3 "Dynamically Updated" Model

The third supplementary model included "dynamic variables" that changed as a function of time. The two dynamic variables used were the number of labs and medications ordered by particular points in time. These variables were assessed from 30 minutes post-ED-arrival until 240 minutes post-ED-arrival, in intervals of 30 minutes, and a distinct model was calculated for every time point for a total of eight models. The performance of each model was evaluated by constructing an ROC curve and calculating the area under the curve (AUC) as outlined in Section 3.9.

3.8 Multivariate Logistic Regression: Model Selection

A number of selection methods were taken to further reduce the number of predictors from 70 to four.

3.8.1 Identification of Early Predictors

Given that the purpose of the model is to predict disposition early in the patient visit, we evaluated which predictors would be available prior to the end of the patient triage step in the ED sequence of events. We identify these variables as the "early predictors."

3.8.2 Lasso Regularization

In order to facilitate further variable reduction, the early predictors were fed into a lasso-regularized logistic regression model [19, 34] using MATLABs lassoglm() function. The function makes a lasso-regularized logistic regression model by using Equation 6 to calculate β. Because the function only accepts numerical matrices as input, all variables (continuous and categorical) were converted to a numeric format. Categorical variables with more than two possible values were converted to binary values with splits as suggested by the univariable results. For example, if Financial Classes A and B were associated with hospital admission while Financial Class C was associated with discharge, then a value of 1 was assigned to Financial Classes A and B and a value of 0 was assigned to Financial Class C. Any patient that had missing data for any of the 29 predictors was removed for this part of the model construction. The predictors having nonzero coefficients at the lambda within one standard deviation of the minimum deviance were then recorded.

3.8.3 Removal of Nonsignificant and Frequently Missing Predictors

Based on the significance results of the univariate regression and frequency with which each predictor was recorded, the number of predictors was further decreased to a manageable size for which to perform a modified "best subset" approach [34].

3.8.4 Model Iteration and Selection

We performed a "best subset" approach to find the four variables that most effectively predicted the outcome. The number four was chosen to ensure an effective model prediction as well as to ensure the model was not too cumbersome for care providers. The performance of each four-variable model was evaluated by using receiver operating characteristic (ROC) analysis and measuring the area under the ROC curve. The model achieving the highest AUC was recorded as the optimal 4-variable model.

3.9 Multivariate Logistic Regression: Model Validation

In order to validate the model we added patient data from the second hospital to that of the first hospital and performed 10-fold cross validation. To measure model performance, predicted values were thresholded at values between 0 and 1, and the sensitivity and specificity were calculated for each threshold. This was used to construct a receiver operating characteristic (ROC) curve. The area under the ROC curve (AUC) was used as the primary measure of model performance, while sensitivity, specificity and accuracy were used as secondary measures. Our AUC value was obtained for the cumulative "test" cohort, not for the training cohort.

CHAPTER IV

RESULTS

4.1 Data and Study Cohort

A total of 5,784 patients were seen in the first ED for primary diagnoses of asthma from January 1st, 2013 to January 31st, 2014. A total of 3,913 patients were seen in the second ED for the same primary reason in between the same start and end times. Baseline and demographic characteristics of these patient cohorts are given in Table 4.1.

Of the 5,784 cases from the first ED, six were excluded because they did not have a listed departure time from the ED, which was necessary to distinguish ED events from inpatient events. An additional 51 cases were excluded because the patients were admitted to the operating room (OR) or transferred to another facility and therefore did not meet clear admission or discharge criteria. After these eliminations, 5,727 cases remained for the subsequent analysis.

Of the 3,913 cases from the second ED, 34 were excluded because the patients were admitted to the OR, transferred to another facility, left against medical advice, or had no disposition recorded. After these eliminations, 3,879 cases remained for subsequent analysis.

4.2 Outcome

Of the 5,727 remaining cases from the first ED, 1,736 patients were admitted (30.3%). Of the 3,879 remaining cases from the second ED, 901 were admitted (23.2%).

Table 6: Baseline Characteristics For Two ED Populations

Variable	**Possible Answers**	**Scottish Rite ED (n = 5784) (%)**	**Egleston ED (n = 3913) (%)**
Gender			
	Male	3575 (61.8)	2427 (62.0)
	Female	2209 (38.2)	1486 (38.0)
Age			
	0 - 18 months	562 (9.7)	150 (3.8)
	18 - 36 months	1048 (18.1)	605 (15.4)
	3 - 6 years	1682 (29.1)	1141 (29.2)
	>6 years	2492 (43.1)	2017 (51.5)
Race			
	American Indian	23 (0.4)	7 (0.2)
	Asian	176 (3.0)	64 (1.6)
	Black	2344 (40.5)	3158 (80.7)
	Hawaiian	5 (0.1)	4 (0.1)
	Other	496 (8.6)	145 (3.7)
	White	2740 (47.4)	535 (13.7)
Triage Acuity			
	ESI 1	3 (0.1)	7 (0.1)
	ESI 2	1516 (26.2)	1067 (27.3)
	ESI 3	2913 (50.4)	2167 (55.4)
	ESI 4	1283 (22.2)	636 (16.2)
	ESI 5	62 (1.1)	33 (0.8)
	Unknown	7 (0.1)	3 (0.1)
ED Disposition			
	Discharge	3995 (69.1)	2978 (76.1)
	Admit to Ward	1598 (27.6)	769 (19.7)
	Admit to ICU	140 (2.4)	132 (3.4)
	Admit to OR	47 (0.8)	17 (0.4)
	Transfer	4 (0.1)	13 (0.3)
	Left AMA	0 (0.0)	2 (0.1)
	Blank/Error	0 (0.0)	2 (0.1)

Table 7: Univariate Regression Results: Demographic Variables

Variable	Possible Answers	OR (95% CI)	*p*
Sex	M=1; F=0 (R)	0.977 (0.87 - 1.09)	0.70
Race			
	White (R)	1	
	Black	0.791 (0.70 - 0.89)	<0.001
	Asian	1.075 (0.78 - 1.48)	0.66
	American Indian	0.715 (0.28 - 1.82)	0.48
	Pacific Islander	0.507 (0.06 - 4.54)	0.54
	Other	0.665 (0.53 - 0.83)	<0.001
Language			
	Spanish (R)	1	
	English	1.914 (1.63 - 2.25)	<0.0001
	15 Other		
Interpreter	Y=1 (R); N=0	1.669 (1.40 - 2.00)	<0.0001

4.3 Univariate Logistic Regression

To ensure that a separate patient cohort remained for model validation, we used only the patients from the first ED for the univariate regression analysis. Univariate analysis was performed for approximately 70 variables.

4.3.1 Demographic Variables

Demographic variables we tested include sex, race, language, and the need for an interpreter. There is no significant relationship between sex and disposition. Compared to Caucasians, African-Americans are admitted to the hospital less frequently (OR = 0.791; p <0.0001). Also admitted less frequently are patients who speak Spanish as a primary language (OR = 1.914 for English speakers; p <0.0001) and patients needing an interpreter (OR = 1.669 for patients not needing an interpreter; p <0.0001).

4.3.2 Financial Variables

Financial variables we tested include primary payor, financial class, and financial class group. It is true that these categories are overlapping; nevertheless, analysis was performed on all of them to see which would be best correlated with outcome.

Table 8: Univariate Regression Results: Financial Variables

Variable	Possible Answers	OR (95% CI)	*p*
Primary Payor			
	Unidentif Priv (R)	1	
	Medicaid	1.562 (1.27 - 1.92)	<0.0001
	Pending Medicaid	11.46 (4.63 - 28.3)	<0.0001
	Generic	2.335 (1.31 - 4.17)	<0.001
	UMR	4.298 (1.52 - 12.1)	<0.001
	21 Other Priv		
Financial Class			
	CMO Medicaid (R)	1	
	Managed Care	1.632 (1.43 - 1.86)	<0.0001
	Medicaid	1.610 (1.35 - 1.92)	<0.0001
	Self-pay	0.543 (0.41 - 0.71)	<0.0001
	Commercial	2.291 (1.34 - 3.91)	<0.001
	Tricare	1.391 (0.85 - 2.29)	0.19
	OOS Medicaid	1.057 (0.59 - 1.88)	0.85
Fin Class Grp			
	Medicaid (R)	1	
	Managed	1.456 (1.29 - 1.64)	<0.0001
	Ind/Char/Self	0.534 (0.42 - 0.68)	<0.0001
	Commerc/Shared	1.552 (1.08 - 2.23)	0.02

Using a private insurer as a reference, Medicaid (OR = 1.562; p <0.0001), pending Medicaid (OR = 11.46; p <0.0001), generic (OR = 2.335; p <0.01), and United Medical Resource (OR = 4.298; p <0.01) patients are significantly more likely to be admitted. Various other private insurers were also used by patients, some of which were significantly related to disposition. More generally, with regard to financial class group, patients who were self-payors (OR = 0.534; p <0.0001) were less likely to be admitted than Medicaid (OR = 1), managed care (OR = 1.456; p <0.0001), or commercial / shared patients (OR = 1.552; p <0.02).

4.3.3 Pre-Arrival/Triage Variables

The IBEX Admit Code variable specifies the patient source; as expected, patients who originated from a healthcare location (e.g. clinic, another ED/hospital) were

significantly more likely to be admitted than those originating from a non-healthcare location (OR = 3.608 22.83; p <0.0001). Patients who originated from another ED or hospital were particularly more likely to be admitted, having odds ratios of 15.30 (p <0.0001) and 22.83 (p <0.0001), respectively. The arrival method is also significantly tied to disposition: patients who arrived by car (OR = 1.630; p <0.0001) or by parent (OR = 1.873; p <0.02) were modestly more likely to be admitted than regular walk-ins, while patients who arrived by hospital-provided means of transportation such as Ground (OR = 18.64; p <0.0001), Air (OR = 29.13; p <0.0001), ambulance (OR = 10.45; p <0.0001), and/or EMS (OR = 8.914; p <0.0001) were extremely more likely to be admitted.

The time of visit also may play a significant role. Patients seen on Friday were more likely to be admitted than any other day of the week (OR = 1.346; p <0.01), and those seen during the late night shift (OR = 1) were more likely to be admitted than those seen during the day (OR = 0.597; p <0.0001) or evening (OR = 0.787; p <0.001).

As expected, the estimated severity of illness at triage, measured by the ESI Index (a scale of one to five, in which one is the most severe) is very significantly predictive of disposition (OR = 0.212 and p <0.0001). A similar variable (IBEX Acuity) measured categorically instead of continuously also showed correlation with admission (OR = 4.338; p <0.0001 for patients admitted as High Risk).

4.3.4 ED Assessment - Vital Sign Variables

Almost all of the six vital signs that we examined, even when taking the earliest values for each vital sign, were predictive of hospital admission. The only variable that was not significantly relevant was temperature; on the other hand, a 1 mmHg increase in systolic blood pressure (OR = 1.008; p <0.001), a 1 mmHg decrease in diastolic blood pressure (OR = 0.992; p <0.01), a 1 resp/min increase in the respiratory rate

Table 9: Univariate Regression Results: Pre-Arrival/Triage Variables

Variable	**Possible Answers**	**OR (95% CI)**	***p***
IBEX Admit Code			
	Non-healthcare (R)	1	
	Phys Office/HMO	3.608 (2.95 - 4.41)	<0.0001
	Other ED	15.30 (9.82 - 23.9)	<0.0001
	Hosp Transfer	22.83 (13.9 - 37.5)	<0.0001
	Outpt Clinic	4.937 (3.08 - 7.89)	<0.0001
	Call Ahead Care	3.113 (1.44 - 6.73)	<0.001
	IC Call Center	inf	
	OOS Hosp Transfer	inf	
ED Adm <72 hrs	Y=1 (R); N=0	0.502 (0.38 - 0.66)	<0.0001
ED Adm >72 hrs	Y=1; N=0 (R)	0.081 (0.03 - 0.17)	<0.0001
Asth Adm <72 hrs	Y=1; N=0 (R)	1.009 (0.57 - 1.76)	0.97
Asth Adm >72 hrs	Y=1; N=0 (R)	0.134 (0.04 - 0.43)	<0.001
IBEX Acuity			
	2+ Resources (R)	1	
	Fast Track	0	<0.0001
	1 Resource	0.013 (0.01 - 0.03)	<0.0001
	High Risk	4.338 (3.81 - 4.93)	<0.0001
	Critical	inf	
ED Visit Shift			
	2300 - 0700 (R)	1	
	0700 - 1500	0.597 (0.51 - 0.69)	<0.0001
	1500 - 2300	0.787 (0.68 - 0.91)	<0.001
Day of Week			
	Sun (R)	1	
	Fri	1.346 (1.09 - 1.65)	<0.001
	Other		NS
IBEX Arrival Method			
	Walk-in (R)	1	
	Ground	18.64 (14.5 - 24.0)	<0.0001
	Car	1.630 (1.35 - 1.96)	<0.0001
	Parent	1.873 (1.15 - 3.04)	0.02
	Ambulance - Other	10.45 (8.40 - 13.0)	<0.0001
	Air	29.13 (8.63 - 98.3)	<0.0001
	Other		
Arrive by EMS	Y=1; N=0 (R)	8.914 (7.73 - 10.3)	<0.0001
Triage Acuity	[1,5]	0.212 (0.19 - 0.24)	<0.0001

Table 10: Univariate Regression Results: Vital Sign Variables

Variable	Possible Answers	OR (95% CI)	*p*
Systolic BP	Continuous	1.008 (1.00 - 1.01)	<0.001
Diastolic BP	Continuous	0.992 (0.99 - 1.00)	<0.01
Temperature	Continuous	0.970 (0.90 - 1.05)	0.42
Resp Rate	Continuous	1.048 (1.04 - 1.05)	<0.0001
Heart Rate	Continuous	1.022 (1.02 - 1.02)	<0.0001
SaO2	Continuous	0.790 (0.77 - 0.81)	<0.0001

Table 11: Univariate Regression Results: Laboratory Variables

Variable	Possible Answers	OR (95% CI)	*p*
Rapid Strep Test	Continuous	0.063 (0.03 - 0.15)	<0.0001
BMP	Continuous	1.647 (1.04 - 2.60)	0.03
CMP	Continuous	0.853 (0.45 - 1.62)	0.63
CBC diff	Continuous	1.095 (0.80 - 1.50)	0.57
Blood Culture	Continuous	0.828 (0.53 - 1.30)	0.41
CRP	Continuous	1.395 (0.94 - 2.06)	0.10
Any Labs	Y=1; N=0 (R)	0.490 (0.37 - 0.64)	<0.0001
Lab Count	Continuous	1.505 (1.45 - 1.57)	<0.0001

(OR = 1.048; p <0.0001), a 1 beat/min increase in the heart rate (OR = 1.022; p <0.0001), and a 1% decrease in oxygen saturation levels (OR = 0.790; p <0.0001) are all associated with an increased chance of admission.

4.3.5 ED Assessment - Laboratory Variables

The rapid strep test, a test given to patients suspected of sore throat infection, was very significantly associated with a discharge disposition (OR = 0.063; p <0.0001). Patients who received a basic metabolic panel (BMP) had a slightly significantly higher chance of being admitted (OR = 1.647; p <0.04). Total number of labs ordered was also associated with hospital admission (OR = 1.505; p <0.0001). The comprehensive metabolic panel (CMP), complete blood count with differential (CBC diff), blood culture, and C-reactive protein (CRP) are not significantly associated with admission.

Table 12: Univariate Regression Results: Radiology Variables

Variable	Possible Answers	OR (95% CI)	*p*
CT Head	Continuous	2.302 (0.58 - 9.22)	0.24
CT Abdomen	Continuous	20.79 (2.63 - 164)	<0.01
CT All	Continuous	6.959 (2.76 - 17.6)	<0.0001
X-ray KUB	Continuous	2.615 (1.36 - 5.04)	<0.0001
X-ray Chest	Continuous	1.577 (1.40 - 1.78)	<0.0001
X-ray All	Continuous	1.743 (1.56 - 1.95)	<0.0001
Rad Count	Continuous	3.356 (3.02 - 3.73)	<0.0001

Table 13: Univariate Regression Results: ED Treatment Variables

Variable	Possible Answers	OR (95% CI)	*p*
IV Antibiotics	Continuous	1.555 (0.99 - 2.44)	0.06
IV Fluids	Continuous	17.37 (14.3 - 21.1)	<0.0001
IV Zofran	Continuous	26.79 (8.26 - 87.0)	<0.0001
Med Count	Continuous	1.272 (1.25 - 1.30)	<0.0001

4.3.6 ED Assessment - Radiology Variables

Almost all of the radiology tests we examined were significantly predictive of admission, with the exception of head CT. Overall CT (OR 6.959; p <0.0001) and radiologic (OR 3.356; p <0.0001) study counts were almost always more significantly associated with admission than individual tests.

4.3.7 ED Treatment Variables

Receiving intravenous (IV) fluids (OR = 17.37 per order; p <0.0001) and IV ondansetron (an anti-nausea medication) (OR = 26.79; p <0.0001) were among the strongest predictor variables that we studied. Medication count was also significant.

4.3.8 Process Variables

From the timestamps of the administrative events we were able to calculate time intervals between events of the visit. Twenty-four distinct time intervals were calculated for each visit. Of these, 21 intervals had a significant predictive value for admission

Table 14: Univariate Regression Results: Process Variables

Variable	**Possible Answers**	**OR (95% CI)**	***p***
ED LOS (hrs)	Continuous	2.448 (2.32 - 2.58)	<0.0001
Arr to TS (min)	Continuous	0.895 (0.88 - 0.91)	<0.0001
Arr to TE	Continuous	0.995 (0.99 - 0.99)	<0.01
Arr to Room	Continuous	0.965 (0.96 - 0.97)	<0.0001
Arr to 1st Prov	Continuous	0.981 (0.98 - 0.98)	<0.0001
Arr to 1st ATN	Continuous	0.981 (0.98 - 0.98)	<0.0001
Arr to Reg	Continuous	1.001 (0.99 - 1.00)	0.13
Arr to Dispo	Continuous	1.008 (1.01 - 1.01)	<0.0001
Arr to Exit	Continuous	1.015 (1.01 - 1.02)	<0.0001
TS to TE	Continuous	1.003 (1.00 - 1.01)	0.04
TE to VSR	Continuous	1.007 (1.01 - 1.01)	<0.0001
TE to Init Assess	Continuous	0.991 (0.99 - 0.99)	<0.01
TE to Room	Continuous	0.965 (0.96 - 0.97)	<0.0001
TE to 1st ATN	Continuous	0.984 (0.98 - 0.99)	<0.0001
TE to 1st Prov	Continuous	0.984 (0.98 - 0.99)	<0.0001
Room to Nurse	Continuous	1.006 (1.00 - 1.01)	<0.001
Room to 1st ATN	Continuous	1.001 (1.00 - 1.00)	0.64
Room to 1st Prov	Continuous	1.001 (1.00 - 1.01)	0.66
Room to 1st Res	Continuous	0.947 (0.93 - 0.96)	<0.0001
Room to Dispo	Continuous	1.011 (1.01 - 1.01)	<0.0001
1st ATN to Dispo	Continuous	1.011 (1.01 - 1.01)	<0.0001
1st Prov to Dispo	Continuous	1.011 (1.01 - 1.01)	<0.0001
1st Prov to Exit	Continuous	1.018 (1.02 - 1.02)	<0.0001
Dispo to Exit	Continuous	1.115 (1.11 - 1.12)	<0.0001

using a criterion of p <0.05. Most notably, the time between patient arrival and triage start (OR = 0.895 per additional minute), arrival and room (OR = 0.965 per additional minute), triage end and room (OR = 0.965 per additional minute) and from making the disposition decision to exiting the ED (OR = 1.115 per additional minute). Many of the time intervals have odds ratios close to one, most likely due to the fact that they are measured on a fine scale.

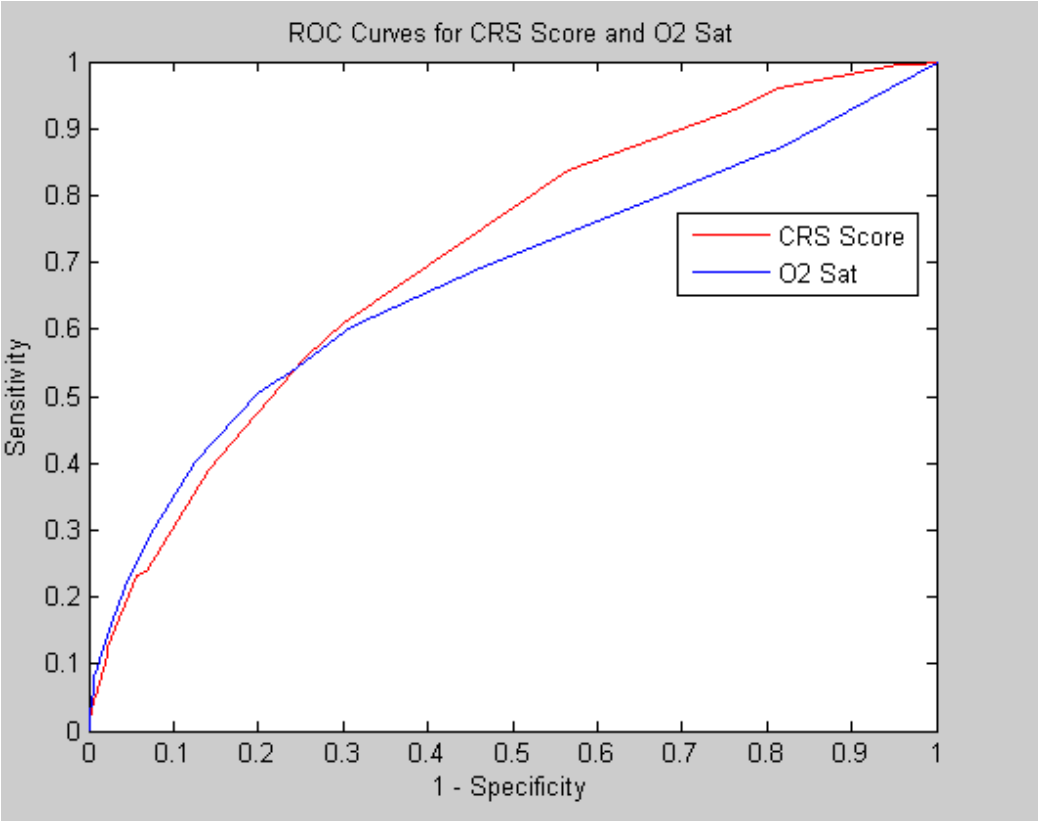

Figure 2: Predictive Abilities of CRS Score and SaO2

4.3.9 One-Variable Prediction: CRS Score vs. O2Sat

To test one of the hypotheses raised in Section 2.5 that objective, easily-obtainable information can predict as well as subjective clinical signs and symptoms, the predictive ability of the CRS score was calculated using univariate logistic regression and compared to that of the oxygen saturation level. Only patients who underwent CRS score testing (approximately 2600 patients) were included in this analysis. ROC curves were constructed (Figure 2). The AUC of the CRS Score model was calculated to be 0.71, with an optimal sensitivity of 61% and specificity of 70%. The AUC of the SaO2 model was almost as good as that of the CRS Score model, calculated as 0.67 with an optimal sensitivity of 60% and specificity of 69%.

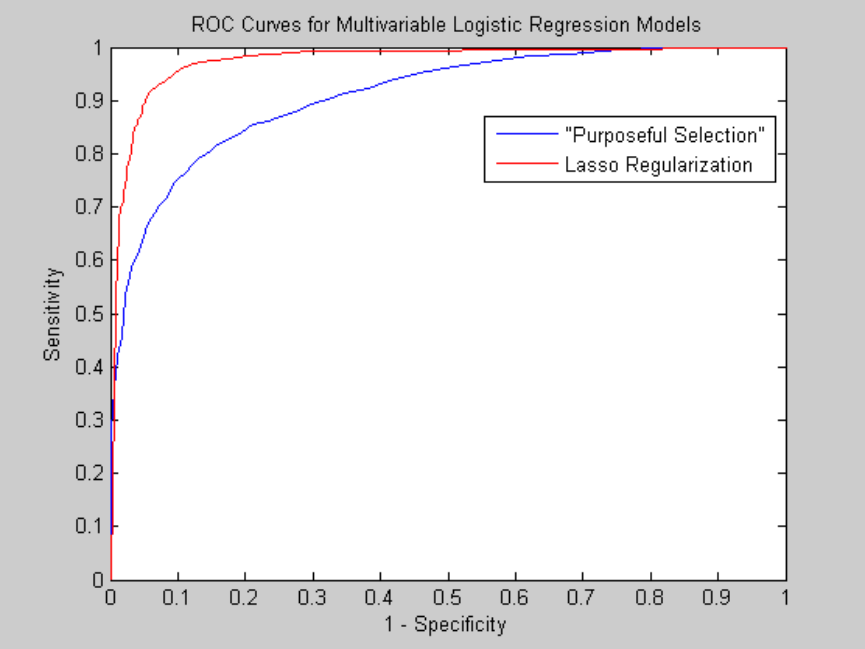

Figure 3: Receiver Operating Characteristic Curves: "Late" Models

4.4 Intermediate Models Containing "Late" Predictors

To support the goals developed in Section 2.5 additional models were made that used "late" variables in addition to "early" predictor variables.

4.4.1 "Purposefully Selected" Model

A model containing predictors selected using a "purposeful selection" approach was constructed (see Table 5 for a list of the included variables). A ROC curve was constructed (Figure 3, blue curve); the area under the curve was calculated to be 0.90, with an optimal sensitivity of 81% and specificity of 82%.

4.4.2 "Lasso Regularized" Model

A model containing predictors selected using a lasso regularization approach was constructed (see Table 5 for a list of selected variables). A ROC curve was constructed (Figure 3, red curve); the area under the curve (AUC) was calculated as 0.97, with an optimal sensitivity of 92% and specificity of 94%.

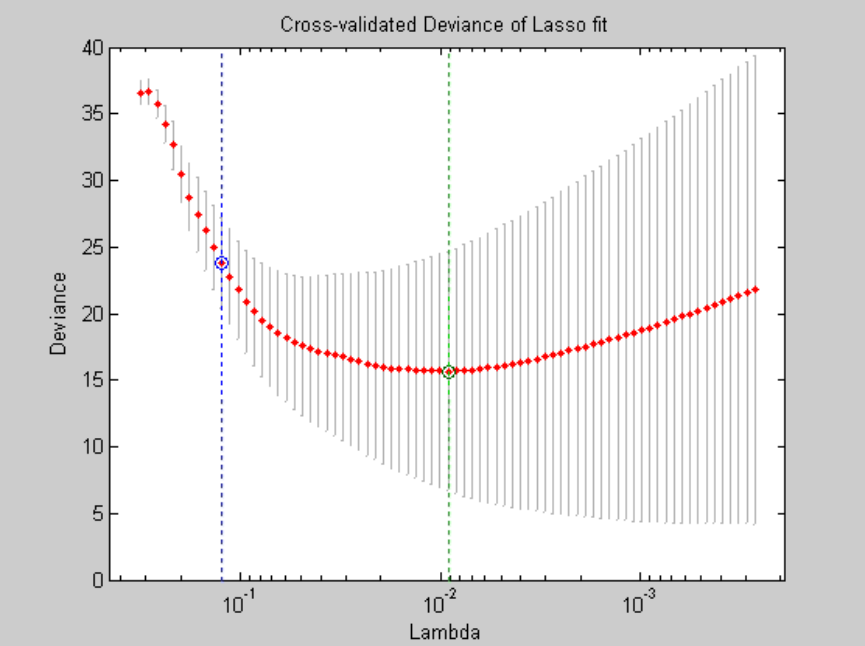

Figure 4: Cross-validated Deviance of Lasso Fit: Late Lasso Regularized Model

4.4.3 "Dynamically Updated" Model

A series of models containing predictors that varied with time was constructed (see Table 5 for a list of included variables). ROC curves were constructed for eight models at 30-minute timepoints, from 30 minutes post-ED-arrival to 240 minutes post-ED-arrival. Four of the ROC curves (at 1-hour, 2-hours, 3-hours, and 4-hours) are shown in Figure 5. The AUCs of these eight models were calculated and were found to improve gradually with time (Figure 6). For the four models shown in Figure 5, the AUCs were 0,855, 0.871, 0.887, and 0.900 at 1, 2, 3, and 4-hours post-arrival, respectively. The optimal sensitivity of the 4-hour model was 82% and the specificity was 81%.

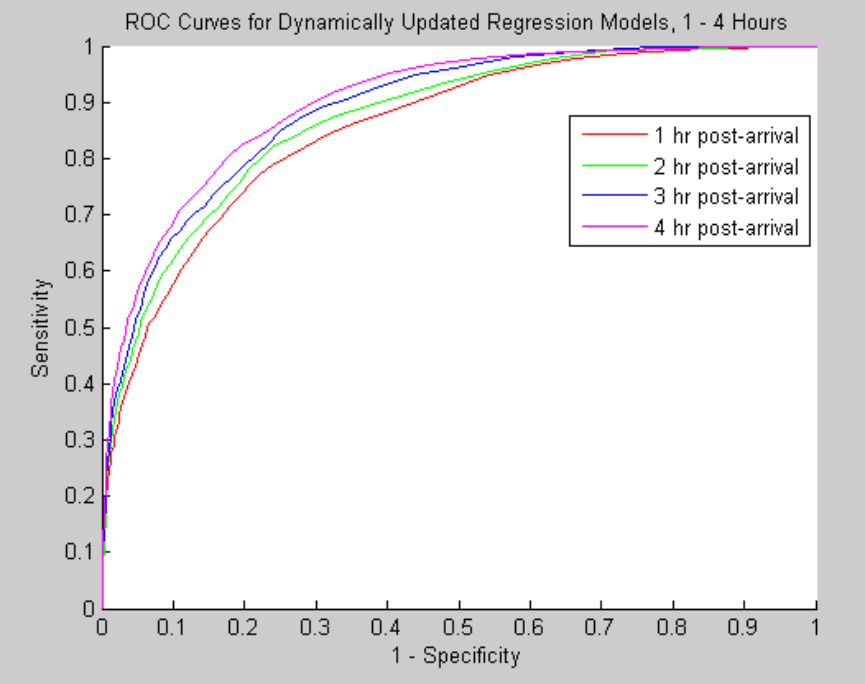

Figure 5: ROC Curves for Dynamically Updated Regression Models: 1 - 4 Hours

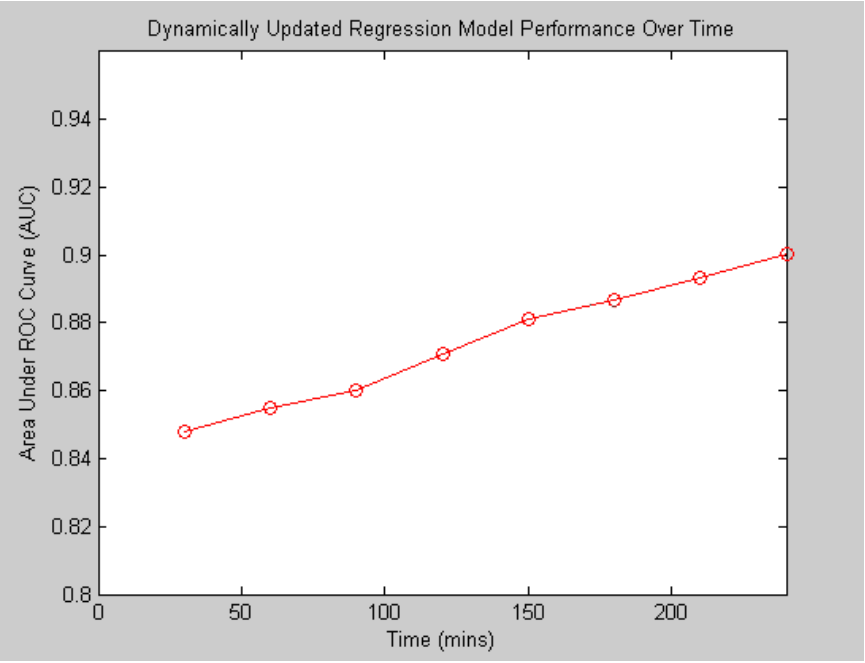

Figure 6: Dynamically Updated Regression Model Performance Over Time

4.5 Multivariate Logistic Regression: Model Selection

4.5.1 Identification of Early Predictors

Twenty-eight of the 70 original predictors were identified as "early predictors" (predictors available prior to the patient entering the ED room for each case). These include all of the demographic variables (Table 4.3.1), all of the financial variables (Table 4.3.2), all of the pre-arrival/triage variables (Table 4.3.3), all of the vital sign variables (Table 4.3.4), and three of the process variables (Table 4.3.8). These 28 variables were fed into the next stage of the main model selection process (the lasso regularization stage, described below).

4.5.2 Lasso Regularization

The 28 "early" predictors were used to develop a lasso-regularized logistic regression model (see Section 3.5.3.2). To determine the optimal λ value, lambda was varied between 1.86×10^{-5} and 1.86×10^{-1} on a 100-point, logarithmic scale. The deviance was calculated for all 100 models, and the model having the lowest deviance within one standard deviation of the minimum deviance was chosen as the optimal model ($\lambda = 0.0072$) (Figure 7, dashed blue line). At this value of λ, the coefficients for 21 of the original 28 predictors were non-zero (Table 4.5.1). The AUC of the model was 0.83, with an optimal sensitivity of 75% and specificity of 74% (Figures 8 and 9). In order to see if elimination of extraneous predictors could improve model performance, the 21 predictors with non-zero coefficients were advanced to the next stage of the optimal model selection process.

4.5.3 Removal of Nonsignificant and Frequently Missing Predictors

In order to reduce the number of possible 4-variable models to a more manageable size for the subsequent "modified best-subset" selection step, five predictors which were nonsignificant in the lasso model and/or the univariate regression analysis were removed from further consideration. In addition, because the systolic and diastolic

Table 15: Model Selection Results: Overview

Variable Name	**Selected by Lasso?**	**Significant and Frequent?**	**Selected by Best Subset?**	**Final Model**
Primary Payor	✓			
Fin Class	✓	✓		
Admission Source	✓	✓	✓	✓
Fin Class Grp				
Sex				
Race	✓	✓		
Language	✓			
Interpreter				
ED Adm <72 hrs	✓	✓		
ED Adm >72 hrs	✓	✓		
Asth Adm <72 hrs				
Asth Adm >72 hrs				
IBEX Acuity				✓
ED Visit Shift	✓	✓		
Day of Week	✓			
IBEX Arrival Method	✓	✓	✓	
IBEX Admit Code	✓	✓		
Arr to TS	✓			
Arr to TE				
TS to TE	✓	✓		
Triage Acuity	✓	✓	✓	
Arrived by EMS	✓	✓		✓
Systolic BP	✓			
Diastolic BP	✓			
Temperature	✓			
Resp Rate	✓	✓		
Heart Rate	✓	✓		
SaO2	✓	✓	✓	✓

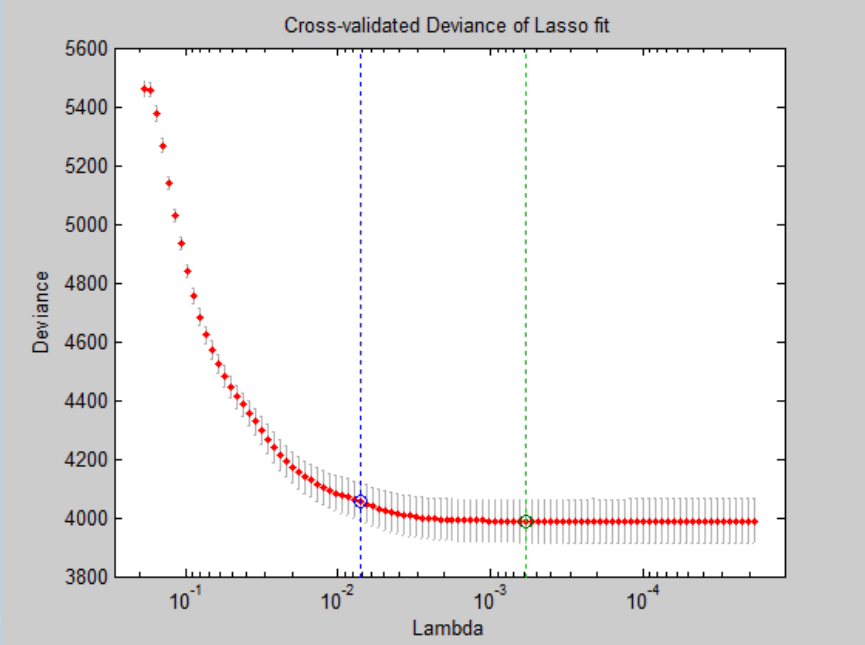

Figure 7: Cross-validated Deviance of Lasso Fit

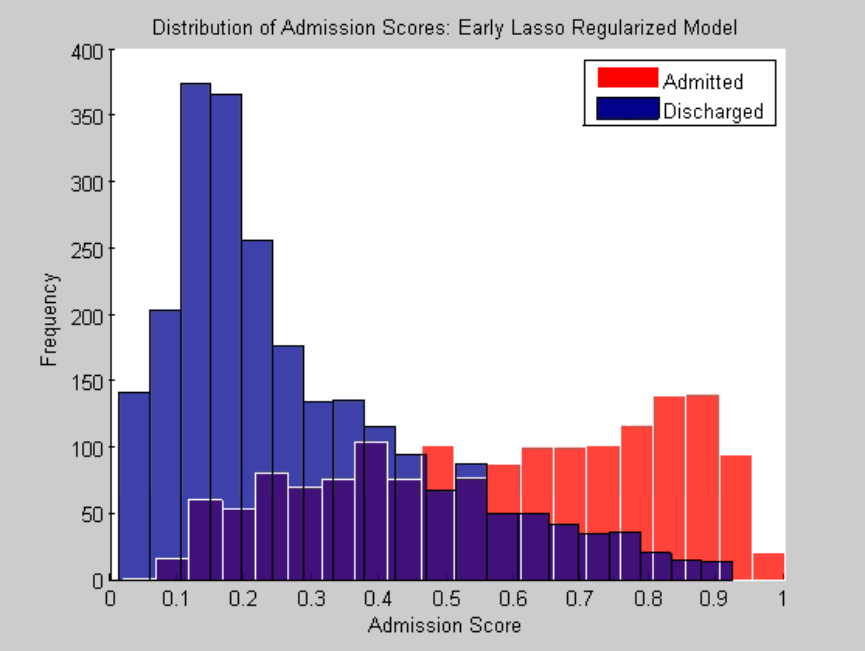

Figure 8: Distribution of Admission Scores: Early Lasso Regularized Model

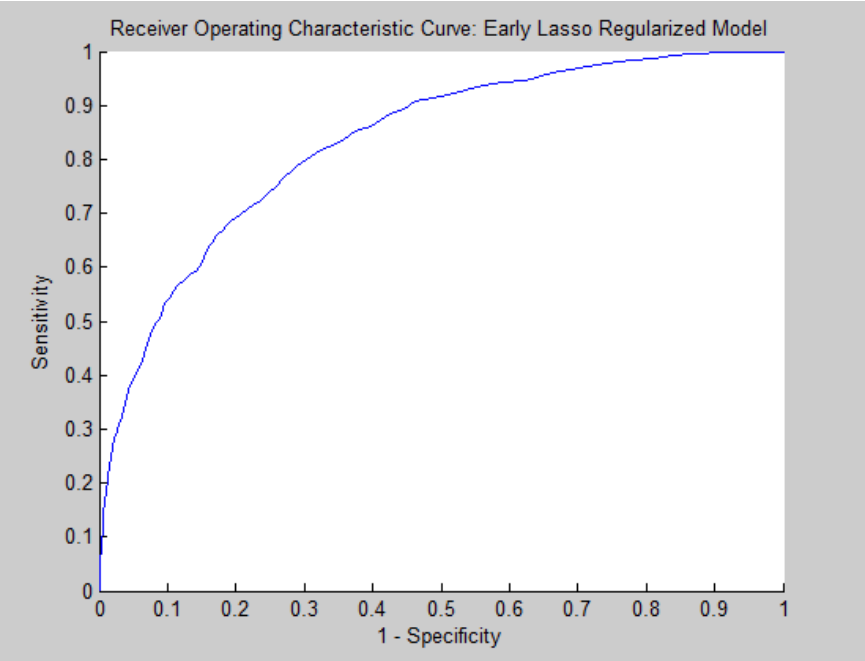

Figure 9: Receiver Operating Characteristic Curve: Early Lasso Regularized Model

blood pressures were missing from a large number of patients these variables were also removed, leaving 14 predictors for the final model selection stage.

4.5.4 Model Iteration and Selection

A "modified best-subset approach" was used to determine the optimal 4-variable model. The best-subset approach used in this study is the same as the traditional best-subset approach, except that the number of predictors is preselected. Of the 1001 total models tested, the 10 four-variable models having the best AUC were recorded (Table 4.5.4). The best model had an AUC of 0.8570 and used the following four predictors: Admission Source, IBEX Arrival Method, Triage Score, and SaO2.

Two adjustments were made to this selected model. The first involved substituting the categorical "IBEX Acuity" variable for the continuous "triage score." Because the former variable was simply a categorical version of the latter, the coefficient of the variable varied non-linearly, ensuring more accurate predictions (AUC = 0.8570 to 0.8662).

The second adjustment substituting the binary categorical "Arrival by EMS" variable for the 30-category "IBEX Arrival Method" variable. This change greatly increased the practicality of the model and its suitability for a web application; instead of the user having to choose between 30 possible values, the user could just make a binary choice. It also increases the generalizability of the model, since the "IBEX Arrival Method" variable contains categories restricted to a particular Metropolitan area. The increased practicality and generalizability came at a very low price, as the AUC was only slightly decreased for the adjusted model (AUC = 0.8662 to 0.8606).

Therefore, the final optimal four-variable model had the following predictors: Admission Source, IBEX Acuity, Arrival by EMS, and SaO2 (Table 4.5.1).

Table 16: Top Ten Early Four-Variable Models: Scottish Rite Hospital

Model #	Financ Cls	Adm Src	Financ Cls Grp	Triage Acuity	IBEX Arriv Meth	IBEX Admit Code	Arr to TS	Arr to TE	TS to TE	Arrive by EMS	RR	HR	SaO2	AUC
1		✓		✓	✓								✓	0.8570
2	✓	✓		✓	✓									0.8523
3				✓	✓	✓							✓	0.8520
4		✓	✓	✓	✓									0.8510
5	✓			✓	✓								✓	0.8508
6		✓		✓						✓			✓	0.8506
7			✓	✓	✓								✓	0.8503
8		✓		✓	✓						✓			0.8482
9				✓	✓							✓	✓	0.8479
10				✓	✓						✓		✓	0.8479

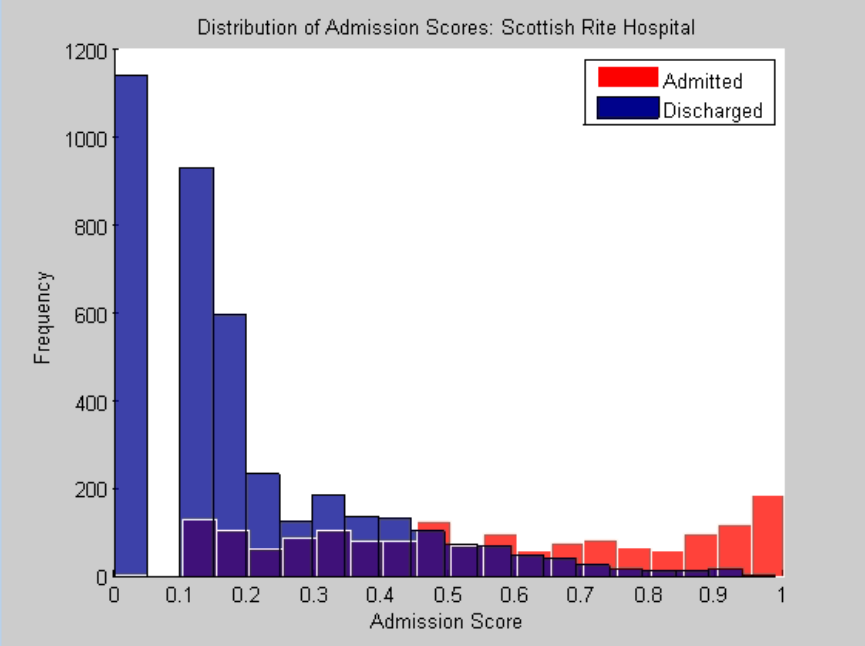

Figure 10: Distribution of Admission Scores: Scottish Rite Hospital

4.6 Multivariate Logistic Regression: Model Validation

The final four-variable model was validated in three ways: 1) by using 10-fold cross validation with the original data used to derive the model; 2) by using 10-fold cross validation with a new data set from a different ED; and 3) by using 10-fold cross validation on the combined data from both EDs.

4.6.1 Metropolitan Hospital 1

Ten-fold cross validation was performed on the original patient sample used to derive the model (n = 5,727). The resulting model had an AUC of 0.859, with an optimal sensitivity and specificity of 77% and 77%, respectively (Figures 10 and 11).

4.6.2 Metropolitan Hospital 2

Ten-fold cross validation was performed using a new, previously unseen patient sample from a different ED (n = 3,879). The resulting model had an AUC of 0.852, with an optimal sensitivity and specificity of 77% and 77%, respectively (Figures 12 and 13).

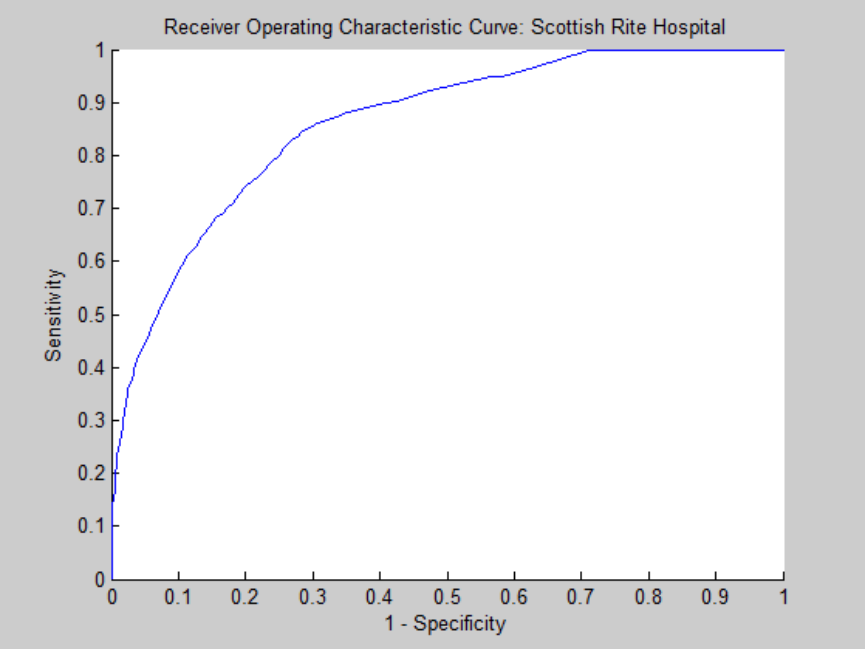

Figure 11: Receiver Operating Characteristic Curve: Scottish Rite Hospital

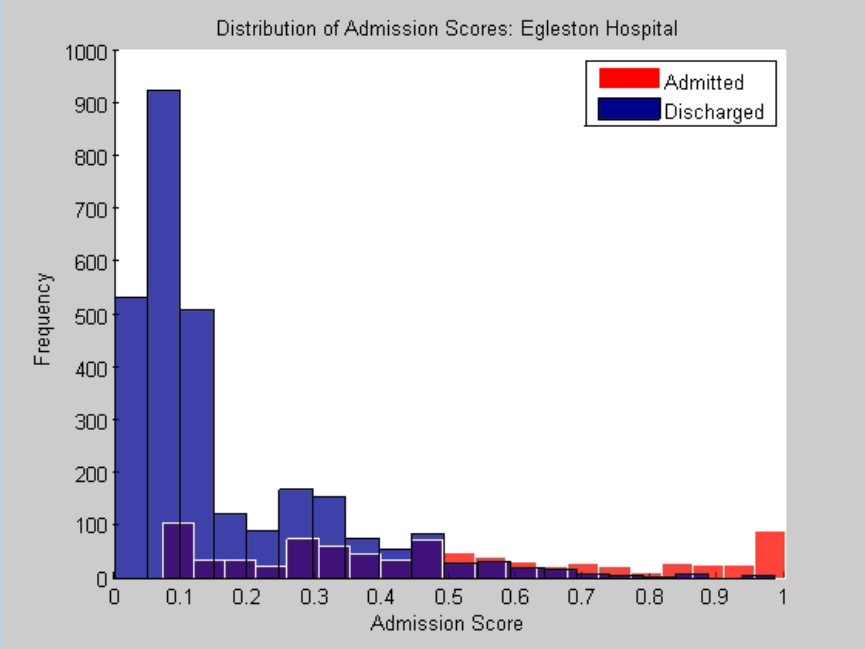

Figure 12: Distribution of Admission Scores: Egleston Hospital

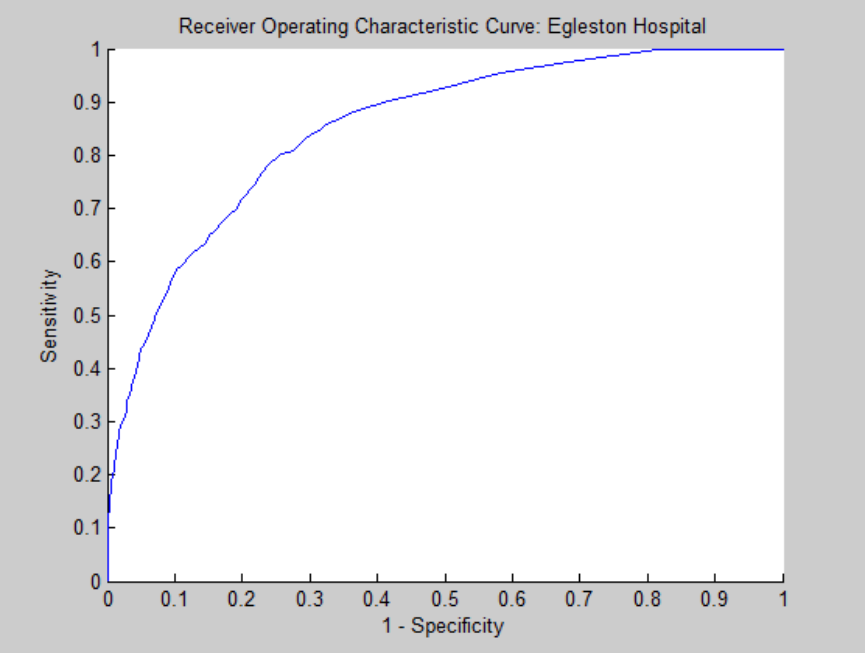

Figure 13: Receiver Operating Characteristic Curve: Egleston Hospital

4.6.3 Metropolitan Hospitals Combined

Finally, 10-fold cross validation was performed using both of the datasets combined (n = 9,606). The resulting model had an AUC of 0.856, with an optimal sensitivity and specificity of 78% and 76%, respectively (Figures 14 and 15).

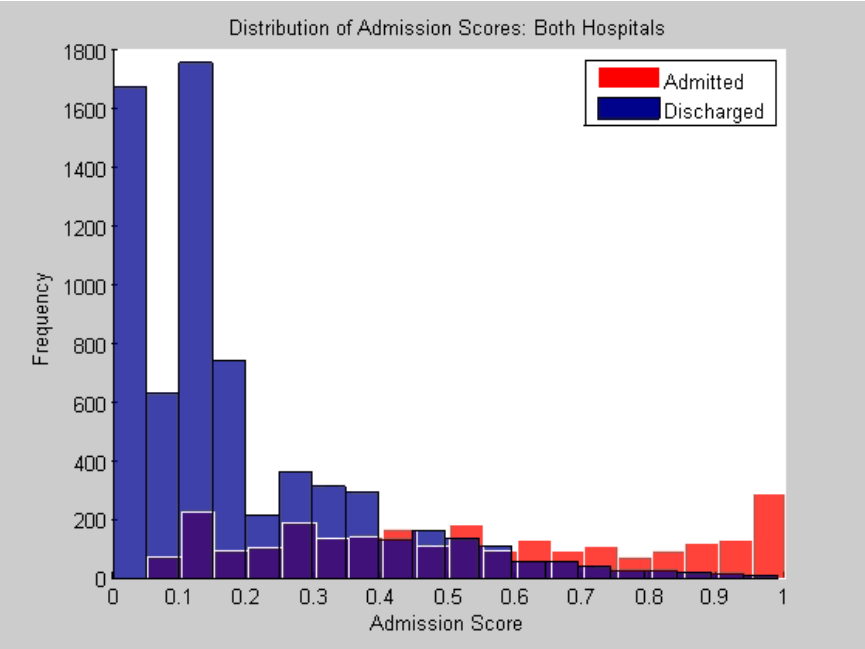

Figure 14: Distribution of Admission Scores: Both Hospitals

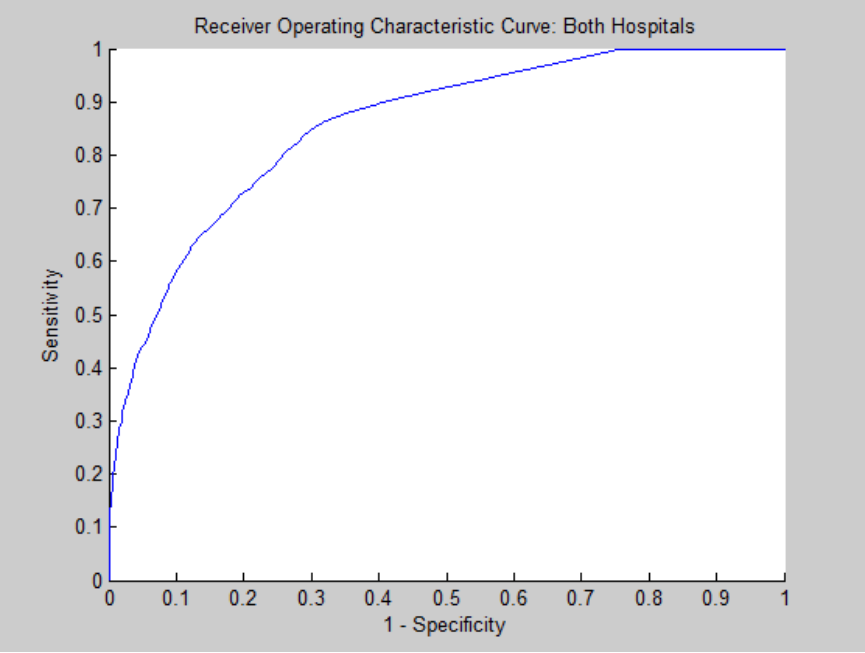

Figure 15: Receiver Operating Characteristic Curve: Both Hospitals

CHAPTER V

DISCUSSION

In this chapter the results are summarized and then discussed in the context of previous work, with an emphasis on novel contributions and significance of this study. Limitations are also reviewed. Finally, implications and future directions for this work are considered.

5.1 Summary of Results

The problem addressed by the study is a traditional machine learning problem in medicine: given some patient's clinical parameters, how can that patient's outcome be predicted? Traditionally, studies attempting to solve this problem use information that is available only late in the patient's visit and/or information that is subjective, unreliable, and difficult to obtain without specialized knowledge, such as measures of symptomatology. Although several recent studies have reasonably overcome these weaknesses for predicting disposition in the emergency department for a general population (see Section 2.2), these studies lacked a tool or set of rules that could be effortlessly used at the point-of-care.

In this study we attempted to address this problem in a subset of ED patients seen for pediatric asthma. For the primary model, a sequence of variable selection methods was used to narrow down the number of predictors from 70 to four; the resulting model achieved excellent predictive performance, having an area under the ROC curve of 0.859 for the training set and 0.852 for the validation set(Sections 4.5 and 4.6). This performance is comparable to that of previously developed asthma prediction models (see Table 2), despite using only four predictors easily obtained early in the patient visit: source of admission, IBEX triage acuity, yes/no arrival by

EMS, and pulse oximetry.

To further facilitate usefulness of this study's findings in the emergency department, a web application is developed based on the primary model (see Chapter 6).

Intermediate models containing "late" predictors (defined as predictors available after the ED physician sees the patient) were also constructed (Section 4.4). It is found that using a lasso-regularized model that includes late predictors results in a drastically improved performance over models using traditional "purposeful selection" techniques (AUC 0.97 vs. 0.90). Finally, construction of models with parameters that vary over time results in increasing performance over time, with a gradual increase in AUC from approximately 0.85 to 0.90 from 1 to 4 hours post-ED arrival.

5.2 Contributions and Significance of Findings

In this section the results of this study are compared to those of previous studies with an emphasis on novel contributions and knowledge gained.

5.2.1 Simple and Early Outcome Prediction

In Section 5.1 the two main pitfalls of most clinical prediction of studies were discussed: 1) the use of information available late in the patient visit, and 2) the use of subjective information requiring specialized knowledge or considerable resources to obtain, such as symptomatologic scores.

The previous asthma prediction study with the best performance [32] demonstrates this point. It had an AUC of 0.89 in the derivation group and 0.92 in the validation group. However, it featured the Pediatric Asthma Severity Score (PASS) as one of its predictor variables. The PASS includes subjective symptomatologic information such as wheezing, work of breathing, and prolonged expiration. It also includes the number of albuterol treatments received in the ED as a predictor, clearly a variable that is not available early in the patient visit.

The supplementary "purposeful selection" model of this study (Section 4.4.1) attempted to rectify one of these two drawbacks by focusing on objective, readily-available information. With an AUC of 0.90, it demonstrated that effective outcome prediction can be achieved without using subjective clinical scores that are difficult to obtain.

The main multivariate model of this study (Sections 4.5 and 4.6) extended this one step further by using only information available before the physician enters the examination room. Despite this, an excellent prediction ability was achieved (AUC 0.855).

5.2.2 Size of Study

The current study surpasses almost all previous studies in the number of subjects. More importantly, it uses an unprecedented amount of predictors (over 70).

5.2.2.1 Number of Subjects

As can be seen in Table 2, previous asthma modeling studies typically contained less than 2,000 patients. This is for two main reasons: 1) most previous studies were prospective, meaning that patients gave permission for inclusion prior to the study, and relevant information was then collected for each subject during ED evaluation and treatment, and 2) most previous studies took place before the 2009 HITECH act which gave incentives to hospitals for having EMR systems. There was one study that examined 2.4 million records [35]; however, multivariate predictive modeling was not a feature of this study.

In contrast, the present study analyzed over 9,500 patient records across two clinical sites in making a multivariate predictive model. Because it was retrospective, no arduous permission seeking or data collection procedures were required. Perhaps future modeling studies could follow this trend and take advantage of increased accuracy and precision of results offered by a larger sample size.

5.2.2.2 Number and Individual Significance of Predictors

This study included univariate regression analysis in its methodology and results. Although not considered a cornerstone of predictive modeling techniques, it is important because of its usefulness in variable selection and model refinement, and also because it provides information as to which predictors are most strongly correlated with the outcome, which may be useful at the point-of-care.

In this study univariate regression was performed on over 70 predictor variables, encompassing a broad number of categories, from demographic and financial data to clinical assessment and treatment data to process data. Seventy represents by far the highest number of predictors examined in any asthma univariate regression study. One previous study [35] examined 2.4 million records but only looked at 12 variables: heart rate, systolic and diastolic blood pressures, respiratory rate, temperature, pulse oximetry, arrival by EMS, sex, triage level, time since last ED visit and hospital discharge, and insurance status. Another previous study [32] examined approximately 29 variables in over 1,200 patients that included demographic data, clinical data such as symptoms and history of asthma severity, care access, baseline and ED albuterol treatment, and ED assessment of pulse oximetry and the PASS score.

By examining such a large number of variables, the present study uncovers new knowledge about previously unused predictor variables and their utility for predicting outcome in asthma patients.

5.2.3 Novel Methodology

The availability of 70 predictor variables, while on one hand is a great asset, can also be a liability when refining a predictive model and making it usable. In this study novel techniques for selecting and pruning these variables are demonstrated.

5.2.3.1 Sequence of Variable Selection Methods

Previous machine learning studies for asthma or other diseases tend to use logistic regression or decision trees as the method for model construction (see Section 2.1.3.2 for a description of these methods and see Table 2 for a summary of methods used in asthma modeling studies). Studies that used logistic regression were faced with the problem of selecting predictor variables to yield the highest-performing model. Two methods were overwhelmingly chosen: 1) a "purposeful selection" technique, in which study authors use their medical knowledge to select predictors that seem like a good fit, often aided with univariate regression results, or 2) a stepwise selection procedure which, as discussed in Section 3.5.3.4, is a greedy algorithm that guarantees neither the optimal subset nor identical subsets across different runs.

In this study an alternative selection procedure was used out of the necessity for eliminating a large number of variables. First, all predictors which became available after the physician saw the patient were eliminated (this can be termed a "modified purposeful selection" step). Forty-two of the 70 predictors were eliminated in this step. Next, a shrinkage technique (the lasso; Section 3.5.3.2) was used to shrink the coefficients of some variables which did not significantly improve performance; seven additional variables were eliminated during this step. In the third stage, a combination of univariate regression and "modified purposeful selection" was used as variables that had frequent missing values or insignificant p-values were removed. In the fourth stage, a "best-subset" approach (Section 3.5.3.1) was used with k (the length of the subset) prechosen as four. Four predictors were used as a compromise between model effectiveness and convenience. Therefore, 1,001 possible models (1,001 is the value of C_4^{14}) were constructed and their performance was analyzed using AUCs; the ten models having the highest AUCs were identified. Finally, in the fifth stage two minor adjustments were made: 1) the replacement of a continuous variable by an otherwise identical categorical variable, and 2) the replacement of a 30-level categorical variable

with a binary categorical variable.

Two of the five stages used ("best-subset" and lasso regularization) are effective ways for variable selection, yet have seldom been used in any healthcare modeling studies. As availability of data continues to grow, techniques such as these will continue to grow in importance, and perhaps the selection procedure used in this study can be duplicated or can serve as a substrate or a stepping stone for the development of alternative procedures.

5.2.3.2 Lasso Regularization in Healthcare Analytics

Aside from the main model of this study, the "supplementary" lasso model (Section 4.4.2) achieved an unprecedented AUC of 0.97. This further demonstrates the usefulness of lasso regularization for model selection.

5.2.4 Demonstration of Improving Performance Over Time

At least two previous studies have addressed the question of how model performance changes over time. In the first study [41], 720 patients were assessed using an Australian clinical asthma scoring system for asthma severity both upon ED arrival and one hour post-arrival. Assessment at 1 hour was more accurate than assessment upon arrival for predicting admission; however, no machine learning methods were implemented. A second study [11] measured the PRAM in up to 297 patients from triage arrival to 4 hours post-arrival using one-hour intervals; they found that assessment at 2, 3, and 4 hours (AUCs 0.85, 0.85, and 0.83, respectively) were better at predicting admission than assessment at triage (AUC 0.76). It should be noted that both of these studies used clinical asthma scoring systems as primary predictors; therefore, the question as to whether objective, readily-available data improves prediction performance over time remains.

The current study addressed this question (Section 4.4.3). Using 12 predictors (see Table 5), two of which (medication count and lab count) changed over time,

models were constructed at 30-minute intervals from 30 minutes post-arrival to four hours post-arrival. A monotonic increase in AUC from 0.855 to 0.900 was observed over one to four hours post-arrival. Therefore, information about clinical symptoms are not needed to improve model performance over time.

5.2.5 Web Application Development

For further details on the web application developed as part of this study, see Chapter 6.

5.3 Limitations

There are many limitations of the current study. Several are noted below:

5.3.1 The "Early" Model as a Clinical Score

While the multivariable models of this study predict outcome very well, they may not perform well in all areas by which clinical scoring systems are evaluated (item selection, reliability, validity, responsiveness, usability, and prediction; see Table 1). In terms of item selection, it can be argued that this study's variable selection method (Sections 3.8 and 4.5) is more sound than the "purposeful selection" method (Section 3.5.3.4). Machine learning methods are also reliable, given that the variables used for the model themselves are objective and reliable. Our models do not use subjective physical examination findings that vary across providers, but rather uses various structured clinical and demographic information that is objective and readily available in EMRs. It therefore is reliable, barring EMR entry and logging errors.

The area of validity is one in which machine learning methods may have a problem, particularly for methods that select variables automatically. For example, in the lasso-regularized model of this study it is unclear why lack of a basic metabolic panel (BMP) being ordered on a patient would be associated with increased admission.

Most machine learning models have not been assessed for responsiveness; that

is, how the value of an observation changes with treatment. For a model to be responsive, some of the variables should be responsive. This means that a variable should be able to show a response to therapy, which may be a problem since there are not many objective measures that show responsiveness to asthma treatment. The oxygen saturation level (SaO2) is one example of this. Spirometry or peak expiratory flow rates would be other objective measures, but these are rarely performed in the ED for many reasons (see Section ??).

Usability is a final domain used to evaluate clinical asthma scores. Therefore, it cannot be concluded that the models developed in this study can serve as a replacement for clinical asthma scores.

5.3.2 Assumptions of Logistic Regression

Logistic regression is a popular choice for machine learning among medical researchers, given its simplicity, effectiveness, and explainability. However, it is a linear method and therefore assumes that the predictor variables can be modeled linearly. [10, 37]. If the variables don't follow this assumption, the effectiveness of the logistic regression can be compromised. Because some of the predictors of this model may not have varied linearly with the response, methods having nonlinear capabilities such as neural networks or polynomial regression may yield more accurate results.

5.3.3 Missing Values

The dataset of this study was not perfect. Several significant predictors had missing values. These ranged from few (the response, triage acuity) to thousands (systolic and diastolic blood pressures). Perhaps imputation techniques would have allowed the consideration of blood pressure measures as viable predictors for the final model.

5.3.4 Binary Response

The response variable for this study was generalized into a binary response (1 = admitted, 0 = discharged). However, this does not make a distinction between patients admitted to the general ward and more severe patients admitted to the ICU. An ordinal or multinomial logistic regression approach would allow prediction of all three outcomes.

5.4 Implications and Future Directions

Based on the limitations discussed in the previous section, there are some ways in which researchers may build on this work. First, alternative methods for dealing with missing data such as imputation could be used. Second, a multinomial model would allow the modeling of a response variable with more than two categories. Finally, nonlinear methods should be explored.

CHAPTER VI

SOFTWARE IMPLEMENTATION

6.1 Previously Developed Web Applications

Clinical decision support can be described as providing health care providers with knowledge and patient-specific information in a timely manner to optimize health and healthcare [9]. Traditional examples of clinical decision support include computerized alert and reminder systems; clinical guidelines; condition-specific order sets; focused data reports and summaries; documentation templates; computer-aided diagnosis and diagnostic support, and contextually relevant reference information [9]. As considerable progress has been made in all of these areas, particularly since the advent of computers at the point-of-care in the 1990s, a full review of clinical decision support is beyond the scope of this work. Interested readers are referred to several existing reviews on this topic [48, 16, 61].

As machine learning methods continue to simmer in the same pot as healthcare data, more and more attention is being paid to computer-assisted prognosis or prediction. Particularly effective prognostic systems would be available at the point-of-care, perhaps using the Internet connected to a computer or mobile device, to guide decision making in real-time.

Since part of the current study is concerned with developing a real-time outcome prediction system for pediatric asthma, in this section known web solutions offering similar outcome prediction capabilities are reviewed.

One website, `www.readmissionscore.org`, contains web programs for predicting 30-day readmission rates for three diseases: heart failure, myocardial infarction, and pneumonia [8]. The applications are targeted for care providers. After selecting the

disease and acknowledging a disclaimer, a screen appears containing 23 patient variables that the provider enters, including information about symptoms, demographics, lab test findings, and EKG test results. Upon clicking "Submit" a percentage is calculated that represents the 30-day readmission risk. Unfortunately all of the 23 variables are required in order to calculate the score. The predictive programs are based on logistic regression models that were constructed using publicly available data released by the Centers for Medicaid Services on over 270,000/560,000/220,000 patients hospitalized with myocardial infarction/congestive heart failure/pneumonia, respectively [45, 40, 46].

A second website focuses on cancer recurrence and survival prediction. It is a project developed at the Memorial Sloan-Kettering Cancer Center (`http://www.mskcc.org/nomograms`) [7]. The homepage lists models for up to 14 different types of cancer including breast cancer, colorectal cancer, lung cancer, and prostate cancer. The models are designed for physicians. Upon clicking "Breast Cancer", an additional page appears with three different outcome prediction models for breast cancer: sentinel lymph node metastasis, additional nodal metastasis, and Ductal Carcinoma in Situ (DCIS) recurrence, all of which are important outcome measures for breast cancer. Each of the three breast cancer models require entry of nine to ten patient variables including age, tumor and lymph node pathological characteristics, hormone status, and other clinical characteristics depending on the model. The outcome is a percentage measure of the likelihood of metastasis/recurrence. Although the webpage states that all of the variables are required, the percentage is reported even if some of the variables are not entered. The breast cancer nomogram for sentinel lymph node metastasis prediction is based on a 2007 study that used logistic regression on 3,786 sentinel lymph node biopsy cases to make a final model having an AUC of 0.754 [17].

A third website (`http://www.predictcancer.org`) focuses on models for estimating survival rates for lung, rectal, head and neck, and endometrial cancer patients

[2]. These models have three to five variables each about functional status, tumor pathology, and patient history depending on the disease. All of the variables require input to obtain a score. Upon submitting the patient information, the percentage probability of two-year survival is returned. A graphical plot is also displayed for survival rates of low, medium, and high risk patients over time; however, this does not change with the input parameters. The lung cancer model is based on a 2009 study that evaluated 377 inoperable, non-small cell lung cancer (NSCLC) patients treated with chemo/radiotherapy. It used Support Vector Machines to obtain an AUC of 0.74 – 0.76 [27].

The final website in this review is more recent and is based on a project at the University of Washington - Tacoma [6]. On the homepage (`http://cwds.uw.edu/health`) is their "Risk-O-Meter" tool that predicts re-hospitalization for heart failure within 30 days. Although the website was down at the time of evaluation, the companion paper [73] states that not all of the ten variables are required for a prediction to be made. In fact only two of the 10 variables (age and gender) are required. After the variables are input, a percentage is calculated that represents the readmission risk. This website differs from the previous websites in a number of aspects: 1) it is designed for patients, not physicians; 2) it uses a Naive Bayes algorithm that is robust to missing data; and 3) it places a greater emphasis on visualization of results.

6.2 Web Application Development

A webpage based on a HTML/CSS/Javascript framework was developed. The D3 Javascript library (D3.js) was used to add visualization touches. The Bootstrap library was used to ensure compatibility with both PCs and mobile devices.

Web forms were used to collect input on the four parameters used in the final model: Admission Source, IBEX Acuity, Arrival by EMS, and oxygen saturation (SaO2). Upon entering all four values and clicking submit, the prediction score is

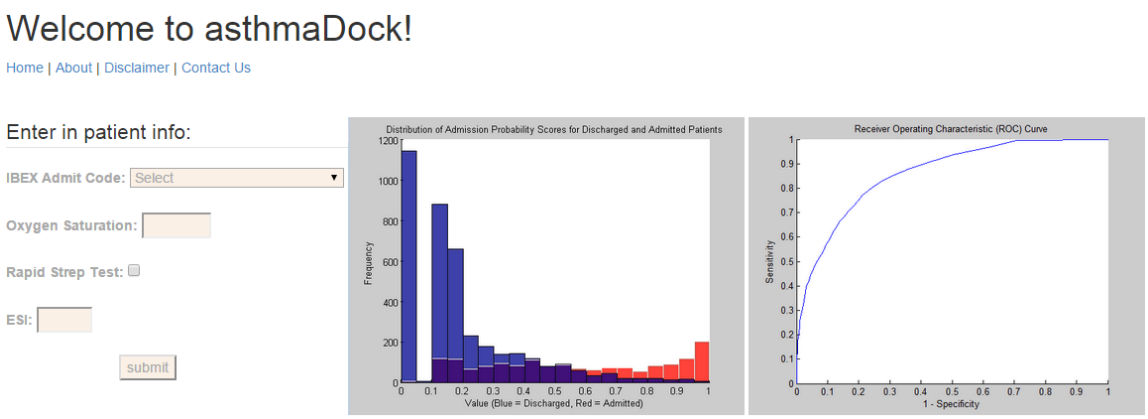

Figure 16: Screenshot of Web Application

displayed for the user. Values toward 1 indicate increased probability of admission. Images of the model histogram and ROC curve for the training set are also included.

6.3 Discussion

The webpage created during this study, like the webpages reviewed in the previous section, is meant to provide prognostic support in healthcare for a specific disease. This webpage is apparently the first that predicts disposition in the emergency department. One limitation is that it could use a visualization and design upgrade.

CHAPTER VII

CONCLUSION

In this study the construction of a practical model that could predict ED outcome in pediatric asthma patients objectively and efficiently was attempted. Furthermore, a web application was developed that offers the possibility of making the model available for predictions at the point-of-care. Some other questions were addressed along the way; for example, how to select variables to be included in the final model; whether the model performance can improve with time; and whether certain methods such as lasso regularization and best-subset have a place in clinical modeling studies. Certainly other researchers can build on this work by internalizing those aspects that worked well and by improving those parts that did not work as well. In conclusion, it is hoped that this study will benefit payors, providers, and patients of the U.S. healthcare system so that the goal of achieving the most effective and efficient care can be achieved.

REFERENCES

[1] "Asthma and Children Fact Sheet." http://www.lung.org/lung-disease/asthma/resources/facts-and-figures/asthma-children-fact-sheet.html. Accessed: 2014-10-22.

[2] "Cancer prediction models by maastro clinic." http://www.predictcancer.org. Accessed: 2015-03-09.

[3] "FastStats - Asthma." http://www.cdc.gov/nchs/fastats/asthma.htm. Accessed: 2014-10-22.

[4] "Generalized Linear Models - MATLAB & Simulink." http://www.mathworks.com/help/stats/generalized-linear-regression.html. Accessed: 2014-11-17.

[5] "Global Health Indicators 2012 - World Health Organization." http://www.who.int/healthinfo/EN_WHS2012_Part3.pdf. Accessed: 2015-02-20.

[6] "Health analytics demos." http://cwds.uw.edu/health. Accessed: 2015-03-09.

[7] "Prediction tools - a tool for doctors and patients." http://http://www.mskcc.org/nomograms. Accessed: 2015-03-09.

[8] "Readmission score calculators." http://www.readmissionscore.org. Accessed: 2015-03-09.

[9] "What is clinical decision support (cds)?." http://www.healthit.gov/policy-researchers-implementers/clinical-decision-support-cds. Accessed: 2015-03-09.

[10] Adler, J., *R in a Nutshell*. Sebastopol, CA: O'Reilly, 2012.

[11] Alnaji, F., Zemek, R., Barrowman, N., and Plint, A., "Pram score as predictor of pediatric asthma hospitalization," *Academic Emergency Medicine*, vol. 21, pp. 872–8, 2014.

[12] Arnold, D., Gebretsadik, T., Abramo, T., Moons, K., Sheller, J., and Hartert, T., "The rad score: a simple acute asthma severity score compares favorably to more complex scores," *Ann Allergy Asthma Immunol*, vol. 107, no. 1, pp. 22–28, 2011.

[13] Baxt, W., "Prospective application of an asthma severity rule," *Acad Emerg Med*, vol. 9, no. 8, p. 868, 2002.

[14] Becker, A., Nelson, N., and Simons, F., "The pulmonary index. assessment of a clinical score for asthma," *Am J Child Dis*, vol. 138, pp. 574–6, 1984.

[15] Bellazzi, R. and Zupan, B., "Predictive data mining in clinical medicine: Current issues and guidelines," *International Journal of Medical Informatics*, vol. 77, pp. 81–97, 2008.

[16] Belle, A., Kon, M., and Najarian, K., "Biomedical informatics for computer-aided decision support systems: A survey," *World Scientific Journal*, vol. 2013, p. 769639, 2013.

[17] Bevilacqua, J., Kattan, M., Fey, J., Cody III, H., Borgen, P., and Zee, K. V., "Doctor, what are my chances of having a positive sentinel node? a validated nomogram for risk estimation," *J Clin Oncol*, vol. 25, pp. 3670–79, year.

[18] Birken, C. S., Parkin, P. C., and Macarthur, C., "Asthma severity scores for preschoolers displayed weaknesses in reliability, validity, and responsiveness," *Journal of Clinical Epidemiology*, vol. 57, pp. 1177–81, 2004.

[19] Bishop, C. M., *Pattern Recognition and Machine Learning*. Springer, 2006.

[20] Braunstein, M. L., *Contemporary Health Informatics*. Chicago, IL: AHIMA Press, 2014.

[21] Cameron, A., Rodgers, K., Ireland, A., Jamdar, R., and McKay, G. A., "A simple tool to predict admission at the time of triage," *Emergency Medical Journal*, vol. 32, pp. 174–9, Mar 2015.

[22] Chalut, D., Ducharma, F., and Davis, G., "The preschool respiratory assessment measure (pram): a responsive index of acute asthma," *J Pediatr*, vol. 137, pp. 762–8, 2000.

[23] Cheng, C.-W., Chanani, N., Venugopalan, J., Maher, K., and Wang, M., "icuarm-an icu clinical decision support system using association rule mining," *Translational Engineering in Health and Medicine, IEEE Journal of*, vol. 1, pp. 4400110–4400110, 2013.

[24] Cinar, M., Engin, M., Engin, E. Z., and Ziya Atesci, Y., "Early prostate cancer diagnosis by using artificial neural networks and support vector machines," *Expert Systems with Applications*, 2009.

[25] Corner, B., "An IOU to Big Pharma Regarding EHR." `http://www.pharmexec.com/iou-big-pharma-regarding-ehr`. Accessed: 2015-02-20.

[26] Dabbous, I., Tkachyk, J., and Stamm, S., "A double blind study on the effects of corticosteroids in the treatment of bronchiolitis," *Pediatrics*, vol. 37, pp. 477–84, 1966.

[27] Dehing-Oberije, C., Yu, S., De Ruysscher, D., Meersschout, S., van Beek, K., Lievens, Y., van Meerbeeck, J., De Neve, W., Rao, B., van der Weide, H., and Lambin, P., "Development and external validation of a prognostic model for 2-year survival of non-small cell lung cancer patients treated with chemoradiotherapy," *Int J Radiat Oncol Biol Phys*, vol. 74, no. 2, pp. 355–62, 2009.

[28] Dexheimer, J., Leegon, J., and Aronsky, D., "Predicting hospital admission at triage in emergency department patients," in *AMIA Annu Symp Proc*, vol. 11, p. 937, 2007.

[29] Ducharma, F., Chalut, D., PLotnick, L., Savdie, C., Kudirka, D., Zhang, X., Meng, L., and McGillivray, D., "The pediatric respiratory assessment measure: A valid clinical score for assessing acute asthma severity from toddlers to teenagers," *J Pediatr*, vol. 152, pp. 476–80, 2008.

[30] Esfandiari, N., Babavalian, M. R., Moghadam, A.-M. E., and Tabar, V. K., "Knowledge discovery in medicine; current issue and future trend," *Expert Systems With Applications*, vol. 41, pp. 4434–4463, 2014.

[31] Farion, K., Michalowski, W., Wilk, S., O'Sullivan, D., and Matwin, S., "A tree-based decision model to support prediction of the severity of asthma exacerbations in children," *J Med Syst*, vol. 34, pp. 551–62, 2010.

[32] Gorelick, M., Scribano, P., Stevens, M., Schultz, T., and Shults, J., "Predicting need for hospitalization in acute pediatric asthma," *Pediatric Emergency Care*, vol. 24, pp. 735–44, Nov 2008.

[33] Gorelick, M., Stevens, M., Schultz, T., and Scribano, P., "Performance of a novel clinical score, the pediatric asthma severity score (pass) in the evaluation of acute asthma," *Academic Emergency Medicine*, vol. 11, pp. 10–18, 2004.

[34] Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning*. Springer, 2009.

[35] Horeczko, T. and Wintemute, G., "Asthma vital signs at triage: Home or admission (ASTHmA)," *Pediatric Emergency Care*, vol. 29, no. 2, p. 175, 2013.

[36] Hosmer, D., Lemeshow, S., and Sturdivant, R., *Applied Logistic Regression*. New York, NY: John Wiley & Sons, 2013.

[37] James, G., Witten, D., Hastie, T., and Tibshirani, R., *An Introduction to Statistical Learning With Applications in R*. New York, NY: Springer, 2013.

[38] Jensen, P., Jensen, L., and Brunak, S., "Mining electronic health records: towards better research applications and clinical care," *Nat Rev Genet*, vol. 13, pp. 395–405, May 2012.

[39] Johnson, J. and Theurer, W., "A stepwise approach to the interpretation of pulmonary function tests," *Am Fam Physician*, vol. 89, no. 5, pp. 359–66, 2014.

[40] Keenan, P., Normand, S., Lin, Z., Drye, E., Bhat, K., Ross, J., Schuur, J., Stauffer, B., Bernheim, S., Epstein, A., Wang, Y., Herrin, J., Chen, J., Federer, J., Mattera, J., Wang, Y., and Krumholz, H., "An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure," *Circ Cardiovasc Qual Outcomes*, vol. 1, pp. 29–37, 2008.

[41] Kelly, A., Kerr, D., and Powell, C., "Is severity assessment after one hour of treatment better for predicting the need for admission in acute asthma?," *Respiratory Medicine*, vol. 98, pp. 777–81, 2004.

[42] Kerem, E., Canny, G., and Tibshirani, R., "Clinical-physiologic correlations in acute asthma in childhood," *Pediatrics*, vol. 87, no. 4, pp. 481–6, 1991.

[43] Kerem, E., Tibshirani, R., Canny, G., Bentur, L., Reisman, J., Schuh, S., Stein, R., and Levison, H., "Predicting the need for hospitalization in children with acute asthma," *Chest*, vol. 98, pp. 1355–61, Dec 1990.

[44] Kononenko, I., "Machine learning for medical diagnosis: history, state of the art and perspective," *Artif Intell Med*, vol. 23, pp. 89–109, Aug 2001.

[45] Krumholz, H., Lin, Z., Drye, E., Desai, M., Han, L., Rapp, M., Mattera, J., and Normand, S., "An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction," *Circ Cardiovasc Qual Outcomes*, vol. 4, pp. 243–52, 2011.

[46] Krumholz, H., Normand, S., Keenan, P., Desai, M., Lin, Z., Drye, E., Bhat, K., and Schreiner, G., "Hospital 30-day pneumonia readmission measure," tech. rep., Yale University/Yale-New Haven Hospital, 2008.

[47] Leegon, J., Jones, J., Lanaghan, K., and Aronsky, D., "Predicting hospital admission in a pediatric emergency department using an artificial neural network," in *AMIA Annual Symp Proc*, p. 1004, 2006.

[48] Madabhushi, A., Agner, S., Basavanhally, A., Doyle, S., and Lee, G., "Computer-aided prognosis: Predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data," *Computerized Medical Imaging and Graphics*, vol. 35, pp. 506–14, 2011.

[49] Mans, R., Schonenberg, H., Leonardi, G., Panzarasa, S., Cavallini, A., Quaglini, S., and van der Aalst, W., "Process mining techniques: an application to stroke care," *Stud Health Technol Inform*, vol. 136, pp. 573–8, 2008.

[50] Mans, R. S., van der Aalst, W. M. P., Vanwersch, R. J. B., and Moleman, A. J., "Process mining in healthcare: Data challenges when answering frequently posed questions," in *Proceedings of the 2012 International Conference on Process Support and Knowledge Representation in Health Care*, BPM' 2012, (Berlin, Heidelberg), pp. 140–153, Springer-Verlag, 2013.

[51] McFadden, E., *Harrison's Principles of Internal Medicine*, ch. Asthma. New York, NY: McGraw-Hill, 2005.

[52] Meisel, Z., Pollack, C., Mechem, C., and Pines, J., "Derivation and internal validation of a rule to predict hospital admission in prehospital patients," *Prehospital Emergency Care*, vol. 12, pp. 314–19, 2008.

[53] Ordonez, C., Omiecinski, E., de Braal, L., Santana, C., Ezquerra, N., Taboada, J., Cooke, D., Krawczynska, E., and Garcia, E., "Mining constrained association rules to predict heart disease," in *Proceedings IEEE Internaitonal Conference on Data Mining 2001*, pp. 433–440, 2001.

[54] Ortiz-Alvarez, O. and Mikrogianakis, A., "Managing the paediatric patient with an acute asthma exacerbation," *Paediatr Child Health*, vol. 17, no. 5, pp. 251–6, 2012.

[55] Ownby, D., Abarzua, J., and Anderson, J., "Attempting to predict hospital admission in acute asthma," *Am J Dis Child*, vol. 138, pp. 162–6, Nov 1984.

[56] Palma, S., Palma, R., Catapani, W., and Waisberg, J., "Predictive factors of hospitalization in children with acute asthma at a university emergency care unit," *Pediatric Emergency Care*, vol. 29, pp. 1175–9, Nov 2013.

[57] Perer, A. and Wang, F., "Frequence: Interactive mining and visualization of temporal frequent event sequences," in *Proceedings of the 19th International Conference on Intelligent Use Interfaces, IUI '14*, (New York, NY), pp. 153–162, 2014.

[58] Piatetsky-Shapiro, G. and Frawley, W., *Knowledge discovery in databases*. California: AAAI/MIT Press, 1991.

[59] Pruikkonen, H., Uhari, M., Dunder, T., Pokka, T., and Renko, M., "Initial oxygen saturation values can predict the need to hospitalise children with mild wheezing," *Acta Paediatrica*, vol. 103, no. 9, pp. 951–6, 2014.

[60] Ramos-Pollan, R., Guevara-Lopez, M., Ortega-Suarez, C., Diaz-Herrero, G., Franco-Valiente, J., and del Solar, M. R., "Discovering mammography-based machine learning classifiers for breast cancer diagnosis," *Journal of Medical Systems*, vol. 36, pp. 2259–69, 2012.

[61] Ritter, S. and Margulies, K., "Emerging tools for computer-aided diagnosis and prognostication," *J Clin Trials*, vol. 4, no. 2, p. e117, 2014.

[62] Rose, O., "Text mining in health records: Classification of text to facilitate information flow and data overview," Master's thesis, Norwegian University of Science and Technology, July 2007.

[63] Schuh, S., Johnson, D., and Stephens, D., "Hospitalization patterns in severe acute asthma in children," *Pediatric Pulmonology*, vol. 23, pp. 184–92, 1997.

[64] Silver, R. and Ginsburg, C., "Early prediction of the need for hospitalization in children with acute asthma," *Clin Pediatr (Phila)*, vol. 23, pp. 81–4, Feb 1984.

[65] Skoner, D., Fischer, T., Gormley, C., Martinez, R., Bobbitt, R., and Holroyde, J., "Pediatric predictive index for hospitalization in acute asthma," *Ann Emerg Med*, vol. 16, pp. 25–31, Jan 1987.

[66] Smith, S., Baty, J., and Hodge, D., "Validation of the pulmonary score: An asthma severity score for children," *Academic Emergency Medicine*, vol. 9, no. 2, pp. 99–104, 2002.

[67] Steyerberg, E., *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY: Springer, 2009.

[68] Sun, Y., Heng, B. H., Tay, S. Y., and Seow, E., "Predicting hospital admissions at emergency department triage using routine administrative data," *Academic Emergency Medicine*, vol. 18, pp. 244–50, 2011.

[69] van der Aalst, W., *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. New York, NY: Springer, 2011.

[70] van der Windt, D., Nagelkerke, A., Bouter, L., Dankert-Roelse, J., and Veerman, A., "Clinical scores for acute asthma in pre-school children: a review of literature," *Journal of Clinical Epidemiology*, vol. 47, pp. 635–46, 1994.

[71] Wilson, J. M. G. and Jungner, G., "Principles and practice of screening for disease," *WHO Chronicle Geneva*, vol. 22, no. 11, p. 473, 1968.

[72] Wright, J. and Young, N., "A comparison of different indices of responsiveness," *Journal of Clinical Epidemiology*, vol. 50, pp. 239–46, 1997.

[73] Zolfaghar, K., Agarwal, J., Sistla, D., Chin, S., Basu Roy, S., and Verbiest, N., "Risk-o-meter: An intelligent clinical risk calculator," in *19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, (Chicago, IL), 2013.